

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA

Programa de Pós-graduação em Biometria

**MODELOS LINEARES GENERALIZADOS
APLICADOS À FILARIOSE BANCROFTIANA**

Fábio Cavalcanti Pereira

DISSERTAÇÃO DE MESTRADO

Recife
22 de fevereiro de 2006

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE ESTATÍSTICA E INFORMÁTICA**

Fábio Cavalcanti Pereira

**MODELOS LINEARES GENERALIZADOS
APLICADOS À FILARIOSE BANCROFTIANA**

Dissertação apresentada ao
**Programa de Pós-Graduação em
Biometria** da Universidade Federal Rural de
Pernambuco, como parte dos requisitos
necessários para a obtenção do grau de
Mestre em Biometria.

Orientador: **Professor Dr. Gauss Moutinho
Cordeiro**, Depto. Estatística e Informática,
UFRPE.

Recife, 22 de fevereiro de 2006

Universidade Federal Rural de Pernambuco
Departamento de Estatística e Informática
Programa de Pós-Graduação em Biometria

Parecer da comissão examinadora da defesa de dissertação de mestrado de

Fábio Cavalcanti Pereira

MODELOS LINEARES GENERALIZADOS
APLICADOS À FILARIOSE BANCROFTIANA

A comissão examinadora, composta pelos professores abaixo, sob a presidência do primeiro, considera o(a) candidato(a) **Fábio Cavalcanti Pereira** como aprovado.

Recife, 22 de fevereiro de 2006

Prof. Dr. Gauss Moutinho Cordeiro - UFRPE
Orientador

Profa. Dra. Jacira Guiro Carvalho da Rocha - UFPE
Membro externo

Prof. Dr. Enivaldo Carvalho da Rocha - UFPE
Membro externo

Prof. Dr. Borko Stosic - UFRPE
Membro interno

À minha família, em especial aos meus pais
Fernando e Nelci, que sempre me deram todo
o suporte e incentivo.

AGRADECIMENTOS

A Deus, pela saúde e força para realização deste trabalho.

Ao meu orientador, Prof. Dr. Gauss Moutinho Cordeiro, pela competência, amizade e principalmente pela confiança depositada em meu trabalho.

Às Profas. Dra. Gerusa Dreyer e Denise Mattos pela forte e valiosa contribuição.

À Profa. Dra. Jacira Guiro, também minha professora desde a graduação, por todo apoio, sugestões e disponibilidade. Por sua amizade, principalmente.

Ao Prof. Dr. Enivaldo Carvalho da Rocha, pela atenção e troca de idéias.

Aos Profs. do Programa de Pós Graduação em Biometria pela atenção e respeito.

A Simara Costa, presente em todos os momentos, pelo apoio, carinho e confiança.

Ao meu tio e amigo Nelson Melo Júnior, por todos ensinamentos ao longo de toda minha vida.

Aos meus avós, em especial a Cecília Regis Cavalcanti, pela paciência e vibração em ouvir minhas histórias e pelos sábios conselhos.

Aos meus irmãos, tios, tias, primos, pela torcida, energia positiva e incentivo.

Aos amigos da Pós-graduação, pela convivência agradável e amizade, em especial a Carlos Batista, Arundo Jr., Leonardo Mendes, Sérgio Paiva, Dâmocles Aurélio, Tony, Eliovânio, Ady Marinho e Franklin.

Aos amigos da Graduação, em especial a Bruno Barbosa, Carlos Gadelha, Catarina Fernandes, Eduardo José, Hugo Assunção, Luis Dimas e Tiago Sales. Pela sinceridade de nossa amizade acima de qualquer outra coisa.

Ao amigo Luciano Souza, pelas palavras de conforto e estímulo.

À CAPES, pelo apoio financeiro concedido.

A todos que, de forma direta ou indireta, contribuíram para execução deste trabalho.

RESUMO

A Filariose Bancroftiana, ocorre como um dos maiores problemas de saúde pública nos trópicos e afeta cerca de 100 milhões de indivíduos. É transmitida por um mosquito e causada pela *Wuchereria bancrofti* que vive nos vasos linfáticos e linfonodos dos seres humanos de todas as idades e ambos os sexos. O substrato anátomo-patológico da doença é a linfangiectasia que leva a uma disfunção linfática produzindo o linfedema (elefantíase), a hidrocele, e as síndromes de fistulização (quilúria, quilocele e linfoescroto). O objetivo dessa investigação se foi estudar a relação entre o risco de desenvolvimento das síndromes fistulizantes e vários parâmetros, tais como a quantidade de gordura na dieta de pacientes atendidos no Núcleo de Ensino, Pesquisa e Assistência em Filariose – NEPAF, Centro de Ciências da Saúde, UFPE. O presente estudo foi realizado usando-se a teoria dos modelos lineares generalizados (MLGs) para ajustar um modelo logístico para os dados da doença usando o programa estatístico S-PLUS.

Palavras-chave: Filariose Bancroftiana, síndromes fistulizantes, modelo logístico, modelos lineares generalizados.

ABSTRACT

Bancroftian filariasis is a major public health problem in the tropics with 100 millions of people infected. It is a vector born infection and it is cause by *Wuchereria bancrofti* a parasite that lives in the human lymphatic vessels and lymph nodes affecting all ages and gender. The basic damage of the lymphatic system is lymphangiectasia which leads to lymphatic dysfunction producing lymphedema (elephantiasis), hydrocece and fistulization syndromes (chyluria, chylocele and lymph scrotum). The aim of this investigation was to study the relationship between the risk of developing fistulation syndrome and several parameters, such a amount of fat on the diet of patients assisted at Center for Teaching, Research and Tertiary Referral for bancroftian filariasis (NEPAF - Federal University of Pernambuco) in Recife, Brazil. The present study was carried out using the generalized linear models (GLMs) for the fitting of logistic model with the statistic program S-PLUS.

Keywords: Bancroftian filariasis; fistulization syndromes, logistic model, generalized linear models.

SUMÁRIO

Introdução	01
Revisão da Literatura	03
Uma Revisão sobre os MLGs.....	03
Formulação do Modelo.....	03
Estimação dos Parâmetros	03
Predição	05
As Componentes de um MLG.....	05
Componente Aleatória.....	05
A Componente Sistemática e a Função de Ligação	09
O Algoritmo de Estimação	10
Verificação da Bondade do Ajuste	14
A Função Desvio	14
A Estatística de Pearson Generalizada X^2	16
Análise do Desvio.....	17
Resíduos	17
Resíduo de Pearson.....	18
Resíduo de Anscombe	18
Desvio Residual.....	19
Filariose Bancroftiana	20
Considerações Gerais sobre a Bacroftose.....	20
Avanços Recentes no Conhecimento da Doença	21

Síndromes de Fistulização	22
Tratamento das Síndromes Fistulizantes	25
Materiais e Método	25
Materiais	25
Método	26
Modelo Binomial	26
Distribuição Binomial	27
Função Geratriz de Momentos e Cumulantes	27
Funções de Ligação Apropriadas	28
A Função de Verossimilhança	32
Estimação dos Parâmetros	33
A Função Desvio	34
Resultados e Discussão	35
Conclusões	41
Referências	43

LISTA DE FIGURAS

Figura 2.1: Quilúria	23
Figura 2.2: Paciente portador de quilocele	24
Figura 2.3: Paciente portador de linfoescroto	24
Figura 4.1: Representação gráfica dos resíduos versus valores ajustados	38
Figura 4.2: Representação gráfica dos resíduos versus percentis da $N(0, 1)$	40

LISTA DE TABELAS

Tabela 2.1: Características de algumas distribuições utilizadas nos MLGs	08
Tabela 2.2: Expressões dos desvios para as principais distribuições da família exponencial	16
Tabela 4.1: Estimativas do parâmetro referente ao modelo logístico com todas as variáveis incluídas	36
Tabela 4.2: Análise do desvio	36
Tabela 4.3: Estimativas dos parâmetros associados ao modelo logístico final	37
Tabela 4.4: Análise do desvio para o modelo logístico final	38

1 INTRODUÇÃO

O modelo de análise de regressão é uma das técnicas mais fortemente usadas em Estatística e há aplicações disto em diferentes áreas. Nesta dissertação consideramos a classe de modelos lineares generalizados (MLGs) desenvolvidos por Nelder e Wedderburn (1972) que desempenham hoje um papel muito importante na Estatística, uma vez que generalizam o modelo clássico de regressão linear, abrindo um leque de opções para a distribuição da variável resposta, permitindo que a mesma pertença à família exponencial de distribuições, bem como dar maior flexibilidade para ligação entre a média e a parte sistemática do modelo. Assim, as hipóteses básicas de normalidade, linearidade e homocedasticidade não são mais exigidas para a análise dos dados.

O modelo linear generalizado é definido por uma distribuição de probabilidade, membro da família exponencial de distribuições, para variável resposta, um conjunto de variáveis independentes descrevendo a estrutura linear do modelo e uma função de ligação entre a média da variável resposta e a estrutura linear.

Várias distribuições de probabilidade importantes (discretas ou contínuas) como normal, gama, poisson, binomial, normal inversa, etc., são membros da família exponencial e os seguintes modelos são casos especiais dos MLGs:

- Modelo logístico;
- Modelo log-linear;
- Modelo probit;
- Modelo normal linear;
- Modelos estruturais com erro gama;

e outros modelos familiares.

Entretanto, os MLGs não englobam dados correlacionados e distribuições fora da família exponencial. Porém, alguns casos especiais de regressão que não são MLGs genuínos podem ser ajustados através de algoritmos iterativos, mediante pequenas alterações, ver Cordeiro e Paula (1992).

A idéia central dos MLGs é transformar as médias dos dados, no lugar de transformar as observações como nos modelos de Box e Cox para se obter um modelo de regressão linear.

A Filariose bancroftiana é uma doença parasitária exclusiva do homem. É causada pela *Wuchereria bancrofti* e transmitida por vetor que, na maioria das regiões endêmicas do mundo, é o *Culex quinquefasciatus*, conhecido no Brasil como muriçoca. A infecção ocorre quando a larva infectante do parasito (L3) penetra no organismo humano após a picada de mosquitos infectados. A L3 ganha o sistema linfático, localizando-se nos vasos linfáticos nos indivíduos adultos e em linfonodos na população pediátrica.

Com os avanços recentes no conhecimento da doença, foi possível verificar que o substrato anatomopatológico da infecção filarial é a dilatação dos vasos linfáticos. A linfangiectasia (dilatação do vaso linfático) causada pelos vermes vivos podem induzir a disfunção linfática, gerando conseqüências diferentes, dependendo da região anatômica em que a mesma esteja localizada. Caso o parasita continue vivo, a linfangiectasia pode progredir potencializando o aparecimento das formas clínicas fistulizantes. O objetivo da presente dissertação é estudar a relação entre pacientes portadores de síndromes fistulizantes e a quantidade de consumo de gordura ingerida na dieta, além de outras variáveis relacionadas ao problema.

Na Seção 2, revisando a literatura, apresentamos a estrutura dos MLGs, o algoritmo de estimação, bondade do ajuste, resíduos e um pouco sobre a filariose bancroftiana. Na Seção 3, apresentamos uma exposição dos materiais e métodos utilizados, a saber: MLGs para análise de dados com resposta binária, com enfoque particular para o modelo logístico. Na Seção 4, ajustamos um modelo logístico para os dados da filariose bancroftiana. Finalmente, na Seção 5, apresentamos as conclusões obtidas.

A plataforma computacional utilizada é o pacote estatístico S-PLUS. O programa S-PLUS é um sistema moderno de manipulação de dados, análise estatística e apresentação gráfica.

2 REVISÃO DA LITERATURA

2.1 Uma Revisão sobre os MLGs

2.1.1 Formulação do Modelo

Nos Modelos Lineares Generalizados, a formulação de um modelo estatístico compreende a escolha de uma distribuição de probabilidade, membro da família exponencial de distribuições, para a variável resposta, de covariáveis para representar a estrutura linear do modelo (matriz modelo), e de uma função de ligação. Para a melhor escolha da referida distribuição de probabilidade para a variável resposta, é aconselhável examinar os dados buscando observar alguns aspectos, tais como: assimetria, natureza discreta ou contínua, intervalo de variação, etc. É importante salientar que os termos que compõem a matriz modelo podem ser de natureza contínua, qualitativa ou mista, e que devem ter uma contribuição significativa na explicação da variável resposta.

Uma importante característica dos MLGs é a suposição de independência, ou pelo menos de não-correlação, entre as observações. Contudo, já existem extensões de MLGs para dados correlacionados. Maiores informações ver Liang e Zeger (1986).

Uma outra característica destes modelos está na distribuição da variável resposta. Considera-se uma distribuição única que deve pertencer à família exponencial. Entretanto, Breslow e Clayton (1993) apresentam extensões de MLGs para modelos de experimentos com efeito aleatório que têm mais de uma componente de erro explícita.

2.1.2 Estimação dos Parâmetros

Uma vez que o modelo tenha sido formulado, o próximo passo é estimar os parâmetros e avaliar as estimativas. No contexto dos MLGs, o processo de estimação é determinado por uma medida (ou critério) de bondade de ajuste entre os dados observados e os valores ajustados gerados a partir do modelo. As estimativas dos parâmetros serão aquelas que minimizam esta medida.

Para obter as estimativas dos parâmetros, deve-se maximizar a verossimilhança, ou log-verossimilhança em relação aos parâmetros, supondo fixo os dados observados. Seja $f(y; \theta)$ a função densidade ou função de probabilidade para a observação y dado o parâmetro θ , então a log-verossimilhança expressa como função do valor médio do parâmetro $\mu = E(Y)$, é dada por

$$l(\mu; y) = \log f(y; \theta)$$

Assim, a log-verossimilhança baseada em uma amostra de observações independentes y_1, \dots, y_n será a soma das contribuições individuais, ou seja,

$$l(\mu; y) = \sum_{i=1}^n \log f_i(y_i; \theta_i),$$

onde $\mu = (\mu_1, \dots, \mu_n)^T$ e $y = (y_1, \dots, y_n)^T$.

McCullagh e Nelder (1983) definem uma medida de bondade do ajuste conhecida como *desvio escalonado* da seguinte forma:

$$D^*(y; \mu) = 2l(y; y) - 2l(\mu; y).$$

Note-se que, para os modelos exponenciais, $l(y; y)$ representa a máxima verossimilhança de um ajuste exato, no qual os valores ajustados são iguais aos valores observados (modelo saturado). Assim, como $l(y; y)$ não depende dos parâmetros de interesse, maximizar $l(\mu; y)$ é equivalente a minimizar $D^*(y; \mu)$ com respeito a μ , sujeito às restrições impostas pelo modelo.

2.1.3 Predição

Esta etapa visa utilizar o modelo estimado para fazer predições. É importante salientar que as quantidades preditas devem estar sempre acompanhadas por medidas de precisão e que o modelo utilizado esteja correto. Para maiores detalhes sobre predições, análise de variância e vários tipos de padronizações, ver Lane e Nelder (1982).

2.2 As componentes de um MLG

A estrutura de um MLG é formada por três partes: uma *componente aleatória* composta de uma variável aleatória Y com n observações independentes, um vetor de médias μ e uma distribuição pertencente à família exponencial; uma *componente sistemática* composta por covariáveis, ou variáveis independentes x_1, \dots, x_p tais que produzam um preditor linear η e uma *função de ligação*, que relaciona as duas componentes citadas acima.

2.2.1 Componente Aleatória

A componente aleatória considera um vetor de observações $y = (y_1, \dots, y_n)^T$ referente às realizações das variáveis aleatórias $Y = (Y_1, \dots, Y_n)^T$, independentes e identicamente distribuídas, com médias $\mu = (\mu_1, \dots, \mu_n)^T$ e pertencentes à família exponencial de distribuições com função de probabilidade dada por

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

onde $a(\cdot)$, $b(\cdot)$ e $c(\cdot)$ são funções conhecidas para cada observação; $\phi > 0$ é denominado *parâmetro de dispersão* e θ é denominado *parâmetro canônico* que caracteriza a distribuição em (2.1). Se ϕ é conhecido, a equação (2.1) representa a família exponencial uniparamétrica indexada por θ .

Seja Y uma variável aleatória com distribuição normal de média μ e variância σ^2 , $Y \sim N(\mu, \sigma^2)$, temos:

$$f_Y(y; \theta, \phi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$$

$$= \exp\left\{\frac{(y\mu - \mu^2/2)}{\sigma^2} - \frac{1}{2}\left(\frac{y^2}{\sigma^2} + \log(2\pi\sigma^2)\right)\right\},$$

onde $\theta = \mu$, $\phi = \sigma^2$, $a(\phi) = \phi$, $b(\theta) = \frac{\theta^2}{2}$ e $c(y; \phi) = -\frac{1}{2}\left\{\frac{y^2}{\phi} + \log(2\pi\phi)\right\}$.

Sendo, $l(\theta, \phi; y) = \log f_Y(y; \theta, \phi)$ a log-verossimilhança para uma única observação, temos uma função de θ e ϕ para um dado y . Assim, através das relações abaixo, podem ser calculadas a média e a variância de Y

$$E\left(\frac{\partial l}{\partial \theta}\right) = 0 \tag{2.2}$$

e

$$E\left(\frac{\partial^2 l}{\partial \theta^2}\right) + E\left(\frac{\partial l}{\partial \theta}\right)^2 = 0 \tag{2.3}$$

Então, a partir de (2.1), temos

$$l(\theta, \phi; y) = \frac{(y\theta - b(\theta))}{a(\phi)} + c(y; \phi),$$

Logo, derivando sucessivamente com relação a θ , temos

$$\frac{\partial l}{\partial \theta} = \frac{\{y - b'(\theta)\}}{a(\phi)} \quad (2.4)$$

e

$$\frac{\partial^2 l}{\partial \theta^2} = -\frac{b''(\theta)}{a(\phi)}. \quad (2.5)$$

Temos, a partir de (2.2) e (2.4), que

$$0 = E\left(\frac{\partial l}{\partial \theta}\right) = \frac{\{\mu - b'(\theta)\}}{a(\phi)}$$

de modo que

$$E(Y) = \mu = b'(\theta). \quad (2.6)$$

Da equação (2.6) podemos obter, univocamente, o parâmetro canônico θ como função da média μ .

Da mesma forma, a partir de (2.3), (2.4) e (2.5), obtemos

$$0 = E\left(\frac{\partial^2 l}{\partial \theta^2}\right) + E\left(\frac{\partial l}{\partial \theta}\right)^2 = -\frac{b''(\theta)}{a(\phi)} + \frac{\text{Var}(Y)}{a^2(\phi)}.$$

Logo,

$$\text{Var}(Y) = b''(\theta)a(\phi). \quad (2.7)$$

Portanto, a variância de Y é o produto de duas funções: $b''(\theta)$ que depende apenas do parâmetro canônico e, por conseguinte, da média, sendo chamada de *função de variância*

$V=V(\mu)$. A *função de variância* expressa como função de μ é reescrita da seguinte forma:

$$V(\mu) = b''(\theta) = \frac{d\mu}{d\theta}. \quad (2.8)$$

E a *função* $a(\phi)$, que depende apenas de ϕ , e é expressa por $a(\phi) = \frac{\phi}{w}$, em que ϕ é o parâmetro de dispersão constante para todas as observações e w é um peso a priori conhecido, que pode variar de observação para observação.

Várias distribuições de probabilidade conhecidas e largamente utilizadas como Normal, Poisson, Binomial, Gama e Normal inversa, são membros da família exponencial de distribuições. Apresentamos na Tabela 2.1 estas distribuições sob a forma (2.1) e algumas de suas principais características (ver McCullagh e Nelder, 1983).

	Notação	ϕ	$b(\theta)$	$c(y; \phi)$	$\mu(\theta) = E(Y; \theta)$	$V(\mu)$
Normal	$N \sim (\mu; \sigma^2)$	σ^2	$\frac{\theta^2}{2}$	$-\frac{y^2}{2\phi}$ $-(\log(2\pi\phi)) / 2$	0	1
Poisson	$P(\mu)$	1	$\exp(\theta)$	$-\log y!$	$\exp(\theta)$	μ
Binomial	$\frac{B(m, \pi)}{m}$	$\frac{1}{m}$	$\log(1 + e^\theta)$	$\log\left(\frac{m}{my}\right)$	$\frac{e^\theta}{(1 + e^\theta)}$	$\frac{\mu}{(1 - \mu)}$
Gama	$G(\mu, \nu)$	ν^{-1}	$-\log(-\theta)$	$\nu \log(\nu y) - \log y$ $-\log \Gamma(\nu)$	$-\frac{1}{\theta}$	μ^2
Normal inversa	$IG(\mu, \sigma^2)$	σ^2	$-(-2\theta)^{\frac{1}{2}}$	$-\frac{1}{2\phi y}$ $-(\log(2\pi\phi y^3)) / 2$	$(-2\theta)^{\frac{1}{2}}$	μ^3

Tabela 2.1: Características de algumas distribuições utilizadas nos MLGs.

2.2.2 A Componente Sistemática e a Função de Ligação

No MLG a componente sistemática, $\eta = (\eta_1, \dots, \eta_n)^T$, também chamada de preditor linear, é uma função linear dos parâmetros desconhecidos $\beta = (\beta_1, \dots, \beta_p)^T$, representada por

$$\eta = X\beta,$$

onde X é uma matriz modelo $n \times p$ ($p < n$) conhecida de posto p . Além disso, outra característica da componente sistemática de um MLG é que a média μ do vetor y é expressa por uma função conhecida (monótona e diferenciável) de η ,

$$\mu_i = g^{-1}(\eta_i), \quad i = 1, \dots, n$$

denominando-se $g(\cdot)$ função de ligação.

Para o modelo clássico de regressão a ligação usual é a identidade ($\eta = \mu$) no sentido de que valores esperados dos dados e preditores lineares podem assumir qualquer valor na reta real $(-\infty, +\infty)$. Entretanto, para modelos com distribuições gama, poisson e normal inversa, a ligação identidade é menos atrativa, uma vez que não restringe os valores esperados ao intervalo $(0, \infty)$. Se Y tem distribuição de Poisson, como $\mu > 0$, note que a função de ligação adequada é a logarítmica ($\eta = \log \mu$), pois esta tem o domínio positivo e o contradomínio na reta real. Para modelos que assumem a distribuição binomial, em que $0 < \mu < 1$, existe a restrição de que o domínio da função de ligação esteja no intervalo $(0, 1)$, enquanto seu contradomínio é o intervalo $(-\infty, +\infty)$. As três principais funções que preservam esta restrição são:

1. logit (ou logística)

$$\eta = \log \left\{ \frac{\mu}{1 - \mu} \right\};$$

2. probit

$$\eta = \Phi^{-1}(\mu),$$

em que $\Phi^{-1}(\cdot)$ é a função distribuição acumulada da normal reduzida;

3. complemento log-log

$$\eta = \log \{ -\log (1 - \mu) \}.$$

2.3 O Algoritmo de Estimação

Para se estimar os valores do vetor de parâmetros β de um MLG, podem ser usados diversos métodos, dentre eles: Bayesiano, qui-quadrado mínimo, estimação – M e o método de máxima verossimilhança. Este último apresenta muitas propriedades ótimas, tais como, consistência e eficiência assintótica, sendo este mais preferido e freqüentemente utilizado nos programas computacionais. O método de máxima verossimilhança será apresentado mais detalhadamente nesta seção.

O algoritmo de cálculo das estimativas de máxima verossimilhança foi desenvolvido por Nelder e Wedderburn (1972) e baseia-se em um método semelhante ao de Newton-Raphson, conhecido como Método Escore de Fisher. A principal diferença em relação ao modelo clássico de regressão é que as equações de máxima verossimilhança são não-lineares.

O método consiste em resolver o sistema $U(\beta)=0$, em que $U(\beta)$ é conhecido como função escore ou função suporte e $l(\beta)$ a log-verossimilhança como função de β

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta},$$

além de utilizar a matriz de informação de Fisher

$$K = \left\{ -E \left(\frac{\partial^2 l(\beta)}{\partial \beta_j \partial \beta_s} \right) \right\} = -E \left(\frac{\partial U(\beta)}{\partial \beta} \right).$$

Expandindo a função escore em série de Taylor até termos de primeira ordem, obtém-se

$$U(\beta^{(m+1)}) = U(\beta^{(m)}) + \frac{\partial U(\beta^{(m)})}{\partial \beta} [\beta^{(m+1)} - \beta^{(m)}] = 0$$

ou ainda

$$\beta^{(m+1)} = \beta^{(m)} - \left[\frac{\partial U(\beta^{(m)})}{\partial \beta} \right]^{-1} U(\beta)^{(m)},$$

onde o índice (m) significa o valor do termo na m-ésima iteração. Este é o método de Newton-Raphson para o cálculo da estimativa de máxima verossimilhança (EMV) $\hat{\beta}$ de β . Já o método escore de Fisher (1925) é obtido pela substituição de $-\frac{\partial U(\beta^{(m)})}{\partial \beta}$ pelo seu valor esperado K .

Considere a componente sistemática, dada por

$$\eta_i = g(\mu_i) = \sum_{r=1}^p x_{ir} \beta_r = x_i^T \beta,$$

onde x_i^T é a i-ésima linha de X .

A log-verossimilhança é dada por

$$l(\beta) = \frac{1}{a(\phi)} \sum_{i=1}^n \{ y_i \theta_i - b(\theta_i) \} + \sum_{i=1}^n c(y_i; \phi).$$

Derivando $l(\beta)$ em relação ao vetor β , tem-se

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \frac{1}{a(\phi)} \sum_{i=1}^n \{y_i - b'(\theta_i)\} \frac{\partial \theta_i}{\partial \beta}.$$

Calculando

$$\frac{\partial \theta_i}{\partial \beta} = \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta}$$

pela regra da cadeia e utilizando as equações (2.6), (2.7) e (2.8), obtemos

$$\mu_i = b'(\theta_i) \text{ e } V(\mu) = b''(\theta_i) = \frac{\partial \mu_i}{\partial \beta_i}.$$

Como x_i^T é a i -ésima linha de X e $\eta_i = x_i^T \beta$, temos

$$\frac{\partial \eta_i}{\partial \beta} = x_i,$$

onde x_i é um vetor coluna $p \times 1$, temos ainda que

$$\frac{\partial \mu_i}{\partial \eta_i} = [g'(\mu_i)]^{-1}.$$

Finalmente, a função escore é expressa por

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \frac{1}{a(\phi)} \sum_{i=1}^n \{y_i - b'(\theta_i)\} \frac{1}{V(\mu_i)g'(\mu_i)} x_i.$$

A matriz de informação para β é dada por

$$K = X^T W X, \quad (2.9)$$

em que W é uma matriz diagonal de pesos definidos por

$$w_i = \frac{1}{a(\phi_i)} V_i^{-1} g'(\mu_i)^{-2}.$$

Então, a função escore usando esta matriz de pesos, é dada por

$$U(\beta) = X^T W z^*,$$

em que z^* é um vetor com dimensão $n \times 1$ expresso por

$$z_i^* = (y_i - \mu_i) \left(\frac{\partial g(\mu_i)}{\partial \mu_i} \right).$$

Com estes dois resultados, o algoritmo escore de Fisher para calcular a EMV de β é expresso por

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

Finalmente, colocando $(X^T W^{(m)} X)^{-1}$ em evidência tem-se,

$$\beta^{(m+1)} = (X^T W^{(m)} X)^{-1} X^T W^{(m)} y^{*(m)}, \quad (2.10)$$

em que $y^{*(m)}$ é uma variável resposta modificada denotada por

$$y^{*(m)} = X \beta^{(m)} + z^{*(m)}.$$

Portanto, a solução das equações de máxima verossimilhança equivale a calcular repetidamente uma regressão linear ponderada da variável dependente modificada y^* sobre X , com matriz de pesos W . Note que, quanto maior for a variância da observação, menor será seu peso no cálculo das estimativas dos parâmetros. Para maiores estudos sobre este algoritmo ver Aitkin et al. (1989).

Os programas computacionais de ajustamento do MLG usam o método score de Fisher para o cálculo das estimativas dos β 's, pois no método de Newton-Raphson existe uma maior probabilidade de não convergência do algoritmo.

2.4 Verificação da Bondade do Ajuste

2.4.1 A Função Desvio

O objetivo principal é a verificação da adequação do modelo como um todo e a realização de um estudo detalhado quanto às discrepâncias locais que, no caso de serem significativas, podem implicar na escolha de um novo modelo. Existem várias medidas para verificar a bondade do ajuste. Uma destas medidas denomina-se *desvio* e equivale à diferença de log-verossimilhanças maximizadas.

Sabemos que ajustar um modelo estatístico a um conjunto de dados é resumir razoavelmente a informação de n observações para p parâmetros, ou seja, é substituir um conjunto de valores observados y por um conjunto de valores ajustados μ , com um número menor de parâmetros. Porém, o modelo mais simples, chamado de modelo nulo, contém apenas um parâmetro que representa a média μ comum a todas observações y 's. Por outro lado, o modelo saturado contém n parâmetros, um para cada observação.

Na prática, procura-se um modelo com p parâmetros situado entre esses dois limites, já que o modelo nulo é muito simples enquanto o modelo saturado é não-informativo. Porém, o modelo saturado é útil para medir a discrepância de um modelo intermediário em investigação com p parâmetros ($p < n$).

Seja $y=(y_1,\dots,y_n)^T$ uma amostra aleatória com distribuição pertencente à família exponencial (2.1). Sejam $\hat{\theta}=\theta(\hat{\mu})$ e $\hat{\theta}=\theta(y)$ as estimativas dos parâmetros canônicos para o modelo em investigação e o modelo saturado, respectivamente. Seja

$$\hat{l}_p = \sum_{i=1}^n l(\hat{\theta}_i, \phi; y_i) = \sum_{i=1}^n \left\{ (y_i \hat{\theta}_i - b(\hat{\theta}_i)) / a_i(\phi) + c(y_i; \phi) \right\}$$

a log-verossimilhança maximizada sobre β para ϕ fixo. Seja

$$\tilde{l}_n = \sum_{i=1}^n l(\tilde{\theta}_i, \phi; y_i) = \sum_{i=1}^n \left\{ (y_i \tilde{\theta}_i - b(\tilde{\theta}_i)) / a_i(\phi) + c(y_i; \phi) \right\}$$

a log-verossimilhança para o modelo saturado com n parâmetros. Assumindo ainda que $a_i(\phi) = \phi / \lambda_i$, podemos escrever

$$2(\tilde{l}_n - \hat{l}_p) = \sum_{i=1}^n 2\lambda_i \{ y_i (\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \} / \phi = D(y; \mu) / \phi = D / \phi .$$

Finalmente,

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

é denominado *desvio* do modelo em investigação, sendo função apenas dos dados e das estimativas de máxima verossimilhança decorrente dos mesmos.

Na Tabela 2.2 apresentam-se as formas da função desvio com $\lambda_i=1$ (caso mais comum) para as principais distribuições da família exponencial. Maiores detalhes são dados por Nelder e Wedderburn (1972).

Modelo	Desvio
Normal	$\sum_{i=1}^n (y_i - \hat{\mu}_i)^2$
Poisson	$2 \sum_{i=1}^n \{y_i \log(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i)\}$
Binomial	$2 \sum_{i=1}^n \{y_i \log(y_i / \hat{\mu}_i) + (m_i - y_i) \log[(m_i - y_i) / (m_i - \hat{\mu}_i)]\}$
Gama	$2 \sum_{i=1}^n \{\log(\hat{\mu}_i / y_i) + (y_i - \hat{\mu}_i) / \hat{\mu}_i\}$
Normal inversa	$\sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / (\hat{\mu}_i^2 y_i)$

Tabela 2.2: Expressões dos desvios para principais distribuições da família exponencial.

2.4.2 A Estatística de Pearson Generalizada X^2

A estatística de Pearson generalizada é uma outra medida importante, a qual é definida da seguinte forma

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

onde $V(\hat{\mu}_i)$ é a função de variância estimada para a distribuição de interesse.

As duas medidas têm, considerando modelo normal linear, distribuição χ^2 exata. Resultados assintóticos são possíveis para outras distribuições. A vantagem da função desvio é que ela é aditiva e que acrescentando-se variáveis explicativas ao modelo, o desvio deve decrescer, diferentemente da estatística de Pearson generalizada.

2.4.3 Análise do Desvio

A análise de desvio, conhecida como ANODEV é uma generalização da análise de variância para os MLGs, com o objetivo de testar modelos encaixados, isto é, cada modelo incluindo mais termos que os anteriores, tendo a mesma ligação e distribuição, visando obter os efeitos de fatores, covariáveis e suas possíveis interações. São considerados modelos encaixados ($M_{pr} < M_{ps}$) quando os termos que formam X_{ps} incluem todos os termos que compõem M_{pr} mais outros termos que não estão em M_{pr} .

Tendo-se uma seqüência de r modelos encaixados $M_{p_1} \subset M_{p_2} \subset \dots \subset M_{p_r}$, com respectivas dimensões $p_1 < p_2 < \dots < p_r$, matrizes $X_{p_1}, X_{p_2}, \dots, X_{p_r}$ e desvios $D_{p_1} > D_{p_2} > \dots > D_{p_r}$, todos eles com a mesma distribuição e função de ligação. É bom ressaltar que as desigualdades entre os desvios não são válidas para a estatística de Pearson generalizada. Em suma, a comparação de modelos encaixados é realizada, exclusivamente, pela função desvio.

As diferenças entre os desvios $D_{p_i} - D_{p_j}$, $p_i < p_j$, devem ser interpretadas como uma medida de variação dos dados, explicada pelos termos que estão em M_{p_j} e não estão em M_{p_i} . Se $D_{p_i} - D_{p_j} > \chi_{p_j-p_i, \alpha}^2$ consideramos que os termos que estão em M_{p_j} e não estão em M_{p_i} são significativos. Para maiores informações, ver McCullagh e Nelder (1983).

2.5 Resíduos

Os resíduos no processo de modelagem estatística têm uma relação muito forte com a qualidade do ajuste feito, constituindo uma das etapas mais importantes no processo de escolha do modelo adequado. Neste sentido, a análise dos resíduos tem uma importância fundamental na verificação da qualidade dos ajustes de modelos. No contexto dos MLGs, os resíduos são usados para explorar a adequação do modelo ajustado com respeito à escolha da função de variância, da função de ligação e de termos no preditor linear. Além

disso, eles também são úteis para indicar a presença de pontos aberrantes, que poderão ser influentes ou não. Os resíduos medem discrepâncias entre os valores observados y_i 's e seus valores ajustados $\hat{\mu}_i$'s.

2.5.1 Resíduo de Pearson

O resíduo de Pearson tem a seguinte expressão:

$$r_{Pi} = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}.$$

A desvantagem deste resíduo é que sua distribuição é, geralmente, bastante assimétrica para modelos não-normais.

2.5.2 Resíduo de Anscombe

Anscombe (1953) propôs uma definição para os resíduos usando uma função $A(y)$ ao invés de y , tal que $A(\cdot)$ é uma função escolhida visando tornar a distribuição de $A(Y)$ próxima à normal reduzida. Barndorff-Nielsen (1978) mostrou, em relação á família exponencial (2.1), que a função $A(\cdot)$ é dada por

$$A(\mu) = \int \frac{d\mu}{V^{1/3}(\mu)}.$$

Portanto, o resíduo de Anscombe visando a normalização e estabilização da variância é expresso por

$$r_{Ai} = \frac{A(y_i) - A(\hat{\mu}_i)}{A'(\hat{\mu}_i) \sqrt{V(\hat{\mu}_i)}}.$$

2.5.3 Desvio Residual

Se D , o desvio, é usado como uma medida de discrepância de um MLG, então, cada unidade de D contribui com uma quantidade

$$d_i = 2\lambda_i \left\{ y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \right\},$$

tal que $\sum_{i=1}^n d_i = D$. Com isso, surge uma nova definição de resíduo, a partir das componentes d_i que formam o desvio, conhecida como desvio residual.

Pregibon (1979) define o desvio residual como

$$r_{Di} = \text{sinal}(y_i - \hat{\mu}_i)\sqrt{d_i},$$

ao invés de d_i pois, segundo ele, se existe uma transformação que venha a normalizar a distribuição do resíduo, então as raízes quadradas das componentes do desvio são resíduos que exibem as mesmas propriedades induzidas por esta transformação. Assim, os resíduos r_{Di} podem ser tratados como variáveis aleatórias tendo aproximadamente distribuição normal reduzida e, conseqüentemente, $r_{Di}^2 = d_i$ tem aproximadamente distribuição χ_1^2 .

No caso do modelo normal, nenhuma distinção é observada entre os três tipos de resíduos. Entretanto, o resíduo de Anscombe e o desvio residual apresentam formas funcionais bem diferentes para modelos não-normais, enquanto que, de uma forma geral, seus valores são bastante próximos para modelos bem ajustados. Já o resíduo de Pearson difere em forma e valor destes dois.

Cordeiro (1986) mostra que os valores de r_A e r_D continuam bastante próximos para os modelos gama e normal inverso. Pierce and Schafer (1986) examinam de forma mais extensiva as definições de resíduos em modelos da família exponencial.

2.6 Filariose Bancroftiana

2.6.1 Considerações Gerais sobre a Bancroftose

A filariose bancroftiana, doença exclusiva do homem, é causada por um parasito intravascular conhecido como *Wuchereria bancrofti*, sendo transmitida por um mosquito que, na maioria das regiões do mundo, é o *Culex quinquefasciatus*, conhecido no Brasil como muriçoca e carapanã. Os parasitos adultos vivem nos vasos linfáticos e linfonodos e após o acasalamento, as fêmeas produzem um grande número de microfilárias (Mf), que atingem o coração direito via ducto torácico, alcançando a circulação geral. No sistema sanguíneo periférico, as Mf são acessíveis ao vetor, por ocasião da hematofagia, reiniciando, dessa forma, o ciclo.

A bancroftose afeta, pelo menos, cerca de 100 milhões de pessoas, distribuídas em 73 países dos diferentes continentes (WHO 1994). A doença de Bancroft é um duro encargo social e econômico inerente aos trópicos e subtropicais da Ásia, da África, do Pacífico Ocidental e de certas regiões das Américas sendo estimado que cerca de 40 milhões de pessoas são portadoras das formas crônicas (Michael et al. 1996). Embora a distribuição da doença pareça global, aproximadamente um terço dos indivíduos infectados reside na Índia, outro terço na África e o restante se encontra, predominantemente, na região ocidental do Pacífico e no sudeste da Ásia. As Américas representam 0,3% da prevalência global e o país de maior número de casos é o Haiti, seguido do Brasil. Em nosso país, são considerados focos de transmissão ativa o Grande Recife, em Pernambuco (Medeiros et al. 1992, Medeiros et al. 1999) com uma estimativa de cerca de 400 mil infectados (André Furtado, comunicação pessoal) e a cidade de Maceió, em Alagoas (Fontes et al. 1998). Belém do Pará, que na década de 50 era a área de maior prevalência, hoje é considerada um foco sob controle. A infecção está geralmente limitada às populações de baixa renda, que vivem em condições sócio-econômicas extremamente precárias.

2.6.2 Avanços Recentes no Conhecimento da Doença

Com a introdução da ultra-sonografia como método para a localização e para a visualização dos vermes adultos vivos de *Wuchereria bancrofti* (Amaral et al. 1994), foi possível verificar que o substrato anatomopatológico da infecção filarial é a dilatação dos vasos linfáticos de causa não obstrutiva, sem qualquer reação inflamatória (Figueredo-Silva et al. 2002). A linfangiectasia (dilatação do vaso linfático) causada pelos vermes vivos podem induzir a disfunção linfática, gerando conseqüências diferentes, dependendo da região anatômica em que a mesma esteja localizada. Quando a linfangiectasia se localiza em pele, isto é, em áreas que têm contato com o exterior (membros superiores e inferiores, mama feminina, parede escrotal ou pênis) pode haver predisposição para as infecções bacterianas secundárias. Em áreas endêmicas, etiologia dos processos agudos é bacteriana em cerca de 97% dos indivíduos acometidos (Dreyer et al. 1999).

A maioria das bactérias causadoras no processo agudo são sensíveis à penicilina. A soma dos edemas (acúmulo anormal de líquido) que acompanham o linfedema (acúmulo de linfa) pode culminar então na sua forma desfigurante chamada de elefantíase. Ao edema se soma o aparecimento de pregueamento e de dobras na superfície do membro. Desse modo, estabelece-se precariedade de higiene, que facilita o acúmulo de bactérias na pele, predispondo o paciente a infecções secundárias que sobrecarregam um sistema linfático já defeituoso. Em uma pequena parcela da população, no entanto, o processo inflamatório agudo é causado diretamente pela morte do verme adulto da filária, denominado de linfangite filarial aguda. Por outro lado, quando o linfático afetado drena linfa de partes internas do organismo, a bactéria não faz parte do processo. Nesse caso, as anormalidades produzidas estão diretamente relacionadas ao parasita filarial. É o caso da hidrocele, quilocele e quilúria (Dreyer et al. 2000).

Assim, resumindo o que foi dito acima, pode-se afirmar que a manifestação clínica, direta ou indiretamente relacionada à infecção filarial, depende da localização do verme adulto, e essa, por sua vez, depende do sexo do indivíduo infectado. Isso é explicado pela localização mais habitual do parasito adulto: nas mulheres, membros inferiores, superiores e mamas

(Dreyer et al. 1996); no homem, linfáticos do cordão espermático constituem a localização mais freqüente para vermes adultos, o que justifica a ocorrência de hidrocele como manifestação clínica prevalente na população masculina (Norões et al. 2003).

2.6.3 Síndromes de Fistulização

Caso o parasita continue vivo, a linfangiectasia pode progredir potencializando o aparecimento das formas clínicas fistulizantes. As síndromes de fistulização são representadas em área endêmica de bancroftose pela quilúria (Figura 2.1), pela quilocele (Figura 2.2) e pelo linfoescroto (Figura 2.3). A ruptura de vasos linfáticos dilatados causa extravasamento de linfa que se acumula em cavidades dando origem a quilocele, como no caso da cavidade vaginal em bolsa escrotal, ou é drenada via fistula por um canal pré-existente, como o sistema excretor urinário, dando origem a quilúria, ou provoca a formação de vesículas linfáticas leitosas na parede de bolsa escrotal, determinado a presença do linfoescroto (ou linfangioma secundário).

A linfa quilosa (quilomicrons) é composta principalmente por gordura em emulsão e fibrinogênio. A gordura da dieta é absorvida pelos linfáticos lácteos do intestino que drenam para a cisterna magna (ducto torácico). Após uma dieta rica em lipídeos a pressão na cisterna do quilo aumenta mais de 10 vezes, contribuindo para uma pressão muito aumentada na parede do vaso linfático. No paciente normal com seus vasos linfáticos íntegros nada deverá acontecer. No caso do paciente portador da doença bancroftiana o sistema linfático por estar incompetente pela linfangiectasia, permite o refluxo de linfa com quilomicrons dando origem as síndromes de fistulização.

Uma dieta normal para um adulto, considerando-se 2000 calorias diárias, é distribuída em 60% de carboidratos, 25% de proteínas e 15% de lipídeos. Nessa distribuição uma pessoa normal poderia consumir até 33g de gordura nas três refeições diárias.

Nas populações mais pobres do planeta, embora as dietas variem de região para região, usualmente o componente lipídico está em proporção além da recomendada, principalmente

pelo custo/benefício na obtenção do valor total calórico para a sobrevivência diária quando o carboidrato fornece 4 quilocalorias por mol, enquanto a gordura, é mais que o dobro, ou seja: 9 quilocalorias por mol. Assim o consumo de lipídeos, está muito acima da proporção de 15% citada acima, substituindo a maioria do componente protéico da dieta pelo seu alto custo e de parte importante dos carboidratos.

A ruptura ou a fistulização de vasos linfáticos dilatados para dentro do sistema excretor urinário leva ao extravasamento de linfa, que se junta à urina, produzindo a quilúria. A sua causa mais comum, em regiões endêmicas, é a filariose bancroftiana, apesar da existência de inúmeras etiologias, como doenças inflamatórias crônicas, por exemplo, a tuberculose, traumatismo abdominal, a gravidez entre outras.



Figura 2.1: *Quilúria* (DREYER G., Mattos D. & NORÕES J. Filariose Bancroftiana In: Dinâmica das Doenças Infecciosas e Parasitárias. José Rodrigues Coura (Ed). Rio de Janeiro. Guanabara Koogan 2005. Cap. 91 pp 1087-1106 (com permissão).

A quilocele é uma doença que acomete a população masculina e pelo imenso grau de inflamação que provoca, é capaz de destruir o testículo. Acomete uma população jovem (com predominância em faixa etária de 20 a 40 anos), engrossando as condições de desemprego, pois o paciente é eliminado no exame médico para admissão. Podem ocorrer em volumes imensos, dificultando sobremaneira o vestuário, a auto-estima, o desempenho sexual e a sua vida social.



Figura 2.2: *Paciente portador de quilocele* (DREYER G., Mattos D. & NORÕES J. Filariose Bancroftiana In: Dinâmica das Doenças Infecciosas e Parasitárias. José Rodrigues Coura (Ed). Rio de Janeiro. Guanabara Koogan 2005. Cap. 91 pp 1087-1106 (com permissão).

O linfoescroto é derivado de uma disfunção linfática grave e complexa, levando ao aumento de pressão retrógrada de vasos linfáticos localizados na derme da bolsa escrotal, produzindo vesículas linfáticas que se exteriorizam e podem romper facilmente. A drenagem da linfa quilosa provoca um grande embaraço para o paciente, por molhar as suas vestes de uma maneira imprevisível. O mal odor característico afeta profundamente os pacientes e a sua vida social.



Figura 2.3: *Paciente portador de linfoescroto* (DREYER G., Mattos D. & NORÕES J. Filariose Bancroftiana In: Dinâmica das Doenças Infecciosas e Parasitárias. José Rodrigues Coura (Ed). Rio de Janeiro. Guanabara Koogan 2005. Cap. 91 pp 1087-1106 (com permissão).

2.6.4 Tratamento das Síndromes Fistulizantes

Todo o paciente com o diagnóstico de infecção ativa (vermes adultos e/ou microfilárias), deve ser tratado com a dietilcarbamzina. Por outro lado, o tratamento antifilarial não reverte o dano linfático sendo necessárias outras medidas de tratamento para melhorar a saúde geral do paciente, assim como especificamente a sua qualidade de vida. Uma dieta pobre em lipídeos e rica em proteína é um dos componentes principais do tratamento para diminuir a pressão intralinfática e conseqüentemente o refluxo de quilomicrons para os linfáticos afetados. O ideal seria que a dieta contivesse, principalmente, triglicerídeos de cadeia média (que são absorvidos pelo sistema venoso e não linfático), o que não é absolutamente factível para os pacientes, pelo seu alto custo.

3 MATERIAIS E MÉTODO

3.1 Materiais

Estudo retrospectivo foi conduzido no Núcleo de Ensino, Pesquisa e Assistência em Filariose – NEPAF do Hospital das Clínicas em Recife/PE, no período de 2003 a 2005. As informações foram extraídas de prontuários dos pacientes atendidos pelo médico assistente. O peso dos pacientes com fístula foi anotado quando os mesmos estavam perdendo linfa (para quilúria e linfoescroto). Para quilocele e os demais pacientes foi considerado o peso do momento do primeiro atendimento.

Critérios de inclusão

- Pacientes de ambos os sexos procedentes do Grande Recife.
- Ter sido atendido pelo Serviço social do NEPAF.
- Ter o seu peso anotado.
- Ter entrevista para quantificação de ingestão de gordura quantificável constando o conteúdo das três principais refeições e de merendas nos intervalos, consumo mensal de óleo e de margarina, manteiga ou banha.
- Ter entrevista da pessoa que prepara o alimento e faz as compras da casa.

- Ter definido o diagnóstico clínico/laboratorial da forma clínica filarial.
- Ter o tempo anotado gasto na entrevista social com os componentes da dieta.

Critérios de exclusão

- Fazer refeições fora de casa.
- Estar abaixo da linha da pobreza ou se alimentando de doação.
- Ter mudança constante de cardápio, tornando impossível a avaliação da ingestão lipídica.
- Ter outras doenças que necessitassem de uma dieta hipolipídica.
- Fazer uso de medicações que impedissem a absorção de lipídeos (tipo Xenical).
- Estar fazendo algum tipo de dieta, independente da doença de base.
- Trabalhar com alimentos ricos em gordura de modo a facilitar uma maior ingestão de lipídios por aumento da oferta (vendedor de coco seco, alimentos como frituras entre outros).

3.2 Método

3.2.1 Modelo Binomial

Em diversas situações a variável resposta de interesse Y tem apenas dois resultados possíveis. Um desses chamado de “sucesso” comumente o resultado mais importante da resposta ou aquele que se pretende relacionar com as demais variáveis e o outro de “fracasso”. É comum encontrar situações práticas em que esse tipo de resposta aparece. Assim, por exemplo: resultado do diagnóstico de um exame laboratorial, positivo ou negativo; resultado de teste de aptidão aplicado a um estudante, aprovado ou reprovado; um paciente pode se recuperar ou não quando submetido a um determinado tratamento, entre outros. Assim, Y_i pode ser definido como igual a 1, se ocorre o sucesso, ou igual a 0 se ocorre o fracasso, atribuindo a Y_i probabilidades de tal forma que:

$$P(Y_i = 0) = 1 - \pi_i ; \quad P(Y_i = 1) = \pi_i , \quad (3.1)$$

onde π_i é a probabilidade de sucesso.

3.2.2 Distribuição binomial

A distribuição binomial é uma das distribuições de probabilidades mais antigas e foi desenvolvida por Bernoulli (1713). Ela surge naturalmente em um grande número de contextos, onde as observações Y são contagens não-negativas limitadas por um valor fixo. Existem duas maneiras de como podemos deduzi-la.

Sendo Y_1 e Y_2 variáveis aleatórias independentes com distribuição de Poisson com médias μ_1 e μ_2 , respectivamente, sabemos que $Y_1 + Y_2$ tem distribuição de Poisson com média $\mu_1 + \mu_2$. Logo, a distribuição condicional de Y_1 dado $Y_1 + Y_2 = m$ é dada por

$$P(Y_1 = y / Y_1 + Y_2 = m) = \binom{m}{y} \pi^y (1 - \pi)^{m-y}, \quad y = 0, 1, \dots, m \quad (3.2)$$

onde $\pi = \mu_1 / (\mu_1 + \mu_2)$.

A notação $Y \sim B(m, \pi)$ denota que Y tem distribuição binomial, expressa em (3.2), com índice m e parâmetro π .

A outra maneira vem da distribuição de Bernoulli, expressa em (3.1), que denota um caso particular da distribuição binomial quando $m=1$. Aqui, temos a distribuição binomial como uma soma $S_m = \sum_{i=1}^m Y_i$ de m variáveis aleatórias Y_1, \dots, Y_m de Bernoulli independentes e identicamente distribuídas.

3.2.3 Função Geratriz de Momentos e Cumulantes

A função geratriz de momentos de (3.1) é

$$M_Y(t) = E\{\exp(tY)\} = 1 - \pi + \pi \exp(t). \quad (3.3)$$

A função geratriz de cumulantes da binomial pode ser facilmente encontrada a partir da soma de funções cumulantes de variáveis aleatórias de Bernoulli independentes.

Então, temos a função geratriz de cumulantes

$$K_Y(t) = \log M_Y(t) = \log\{1 - \pi + \pi \exp(t)\}.$$

Portanto, a função geratriz de momentos da soma estocástica $S_m = Y_1 + \dots + Y_m$ é

$$M_{S_m}(t) = \{1 - \pi + \pi \exp(t)\}^m$$

e sua função geratriz de cumulantes correspondente iguala

$$\log M_{S_m}(t) = m \log\{1 - \pi + \pi \exp(t)\}. \quad (3.4)$$

Finalmente, expandindo (3.4) em série de Taylor e avaliando no ponto $t = 0$, encontramos os quatro primeiros cumulantes da distribuição binomial expressos por

$$K_1 = m\pi;$$

$$K_2 = m\pi(1 - \pi);$$

$$K_3 = m\pi(1 - \pi)(1 - 2\pi);$$

$$K_4 = m\pi(1 - \pi)\{1 - 6\pi(1 - \pi)\}.$$

3.2.4 Funções de Ligação Apropriadas

Para investigar a relação entre a probabilidade de sucesso π da variável resposta e o vetor de covariáveis (x_1, \dots, x_p) assumimos que a dependência entre π e (x_1, \dots, x_p) ocorre através da combinação linear

$$\eta = \sum_{j=1}^p \beta_j x_j.$$

Entretanto, como $-\infty < \eta < \infty$, expressar π através de uma função linear de η seria errôneo do ponto de vista probabilístico, pois π não ficaria restrito ao intervalo $(0, 1)$. Então, uma maneira simples e eficaz de solucionar este problema é o uso de uma transformação $g(\pi)$ que relacione o intervalo unitário à reta real, de tal forma que

$$g(\pi_i) = \eta_i = \sum_{j=1}^p x_{ij} \beta_j; \quad i = 1, \dots, n.$$

As três principais funções de ligação usualmente empregadas no modelo binomial, na qual garantem as restrições sobre a probabilidade π são:

I. Logit ou função logística

$$g_1(\pi) = \log \left\{ \frac{\pi}{1-\pi} \right\};$$

II. Função probit ou inversa da função acumulada da normal reduzida

$$g_2(\pi) = \Phi^{-1}(\pi);$$

III. Complemento log-log

$$g_3(\pi) = \log \left\{ -\log(1-\pi) \right\}.$$

As três funções: logística, probit e complemento log-log, possuem inversa, são contínuas e crescentes no intervalo $(0, 1)$. A função logística possui algumas características que a tornam preferida em relação às outras:

- Pode ser interpretada como o logaritmo da razão de chances;
- Apresenta propriedades mais simples;
- É mais conveniente para análise de dados coletados de forma retrospectiva.

A ligação logística é bastante empregada em estudos toxicológicos e epidemiológicos. Contudo, isto não quer dizer que as outras transformações não são utilizadas na prática. Bliss (1935) iniciou a modelagem de proporções, utilizando um modelo binomial com ligação probit. Já a complemento log-log é recomendada por Collet (1994) quando a distribuição das proporções é bem assimétrica.

Para se compreender melhor o ajuste obtido é necessário a utilização da relação entre π e o preditor linear $\eta = X\beta$. A ligação logística satisfaz

$$\log \left\{ \frac{\pi}{1-\pi} \right\} = \eta = X\beta .$$

Expressando-a em termos do preditor linear, temos

$$\pi = \frac{\exp(\eta)}{1 + \exp(\eta)} .$$

Logo, se a parte sistemática do modelo para uma determinada observação tende para um valor muito negativo, sua probabilidade de sucesso tende para zero. Por outro lado, se a mesma tende para um valor muito grande, esta probabilidade tende para um.

Similarmente, pode-se calcular a relação entre π e η para as outras duas ligações:

$$\pi = g_2^{-1}(\eta) = \Phi(\eta)$$

e

$$\pi = g_3^{-1}(\eta) = 1 - \exp\{-\exp(-\eta)\} .$$

Outras duas famílias de transformações para dados binários foram propostas por Aranda-Ordaz (1981). A primeira é dada por

$$T_{\lambda}(\pi) = \frac{2\pi^{\lambda} - (1-\pi)^{\lambda}}{\lambda\pi^{\lambda} + (1-\pi)^{\lambda}}, \quad (3.5)$$

onde π denota a probabilidade de sucesso e λ representa o parâmetro da transformação.

Dois características importantes de (3.5) são $T_{\lambda}(\pi) = -T_{\lambda}(1-\pi)$ e $T_{\lambda}(\pi) = -T_{-\lambda}(\pi)$, ou seja, T_{λ} trata sucesso e fracasso de forma simétrica. A família $T_{\lambda}(\pi)$, é chamada de simétrica. Para $\lambda=0$, (3.5) se reduz à transformação logística, enquanto que para $\lambda=1$ obtemos a transformação linear.

Já em situações onde é apropriado tratar sucesso e fracasso de forma assimétrica (Yates, 1955), uma segunda família de transformações é proposta. Tal família é definida por

$$W_{\lambda}(\pi) = \frac{\{(1-\pi)^{-\lambda} - 1\}}{\lambda}. \quad (3.6)$$

Aqui, assumimos que

$$\log W_{\lambda}(\pi) = \eta,$$

onde η é real e assume uma expressão linear da forma $\eta = X\beta$.

A expressão (3.6) se reduz a transformação logística quando $\lambda=1$ e a complemento log-log quando $\lambda=0$. Para maiores detalhes sobre estas famílias de ligações, ver Aranda-Ordaz (1981). Além disso, é importante ressaltar o trabalho de Stukel (1988) que generaliza o trabalho apresentado por Aranda-Ordaz (1981).

3.2.5 A Função de Verossimilhança

Considerando os dados y_1, \dots, y_n como valores observados de variáveis aleatórias independentes Y_1, \dots, Y_n com distribuição binomial de índice m_i e parâmetro π_i , respectivamente, temos, a partir de (3.2), que a log-verossimilhança de π dado y é escrita da seguinte forma

$$l(\pi; y) = \sum_{i=1}^n \left[y_i \log \left(\frac{\pi_i}{1 - \pi_i} \right) + m_i \log (1 - \pi_i) \right]. \quad (3.7)$$

O termo $\sum \log \binom{m_i}{y_i}$ pode ser omitido, pois não envolve o parâmetro π .

A log-verossimilhança também pode ser escrita em função do preditor linear. Para isso é necessário a utilização da equação

$$g(\pi_i) = \eta_i = \sum_{j=1}^p x_{ij} \beta_j, \quad i = 1, \dots, n.$$

Se a função escolhida para o modelo for a logística, obtém-se

$$g(\pi_i) = \eta_i = \log \left\{ \frac{\pi_i}{1 - \pi_i} \right\} = \sum_{j=1}^p x_{ij} \beta_j, \quad i = 1, \dots, n.$$

Expressando a log-verossimilhança em função dos parâmetros desconhecidos, temos

$$l(\beta; y) = \sum_{i=1}^n \sum_{j=1}^p y_i x_{ij} \beta_j - \sum_{i=1}^n m_i \log \left\{ 1 + \exp \left(\sum_{j=1}^p x_{ij} \beta_j \right) \right\}.$$

Um ponto importante que deve ser ressaltado é que a estatística $X^T y$, que aparece na log-verossimilhança, é suficiente para β , pois a ligação logística também é a ligação canônica no modelo binomial.

3.2.6 Estimação dos Parâmetros

Para estimarmos os parâmetros usando o método escore de Fisher, apresentado na Seção 2.3, basta calcular a função escore e a matriz de informação de Fisher para a log-verossimilhança do modelo binomial, obtendo-se

$$U(\beta) = X^T (y - \mu)$$

e

$$K = X^T W X,$$

onde

$$W = \text{diag} \{ m_i \pi_i (1 - \pi_i) \}.$$

Finalmente, o algoritmo de estimação de β é dado por

$$\beta^{(m+1)} = \beta^{(m)} + K^{(m)-1} U(\beta^{(m)}).$$

É importante salientar que neste algoritmo as observações com maior variância

$$V(\pi_i) = m_i \pi_i (1 - \pi_i),$$

tem menor peso w_i para o cálculo da estimativa do vetor β .

3.2.7 A Função Desvio

Sabemos que a função desvio corresponde a duas vezes a diferença entre as log-verossimilhanças maximizadas, sob o modelo saturado e sob o modelo em investigação. Sob o modelo em investigação, com probabilidade estimada $\hat{\pi}$, a log-verossimilhança é dada por

$$l(\hat{\pi}; y) = \sum_i \{ y_i \log \hat{\pi}_i + (m_i - y_i) \log (1 - \hat{\pi}_i) \},$$

onde $\hat{\pi}_i = \pi(\hat{\mu}_i) = \hat{\mu}_i / m_i$.

No modelo saturado, a EMV de π_i é obtida por $\tilde{\pi}_i = y_i / m_i$.

Assim, a função desvio para o modelo binomial é expressa como

$$D(y; \hat{\pi}) = 2l(\tilde{\pi}; y) - 2l(\hat{\pi}; y) =$$
$$2 \sum_i \left\{ y_i \log (y_i / \hat{\mu}_i) + (m_i - y_i) \log \left(\frac{m_i - y_i}{m_i - \hat{\mu}_i} \right) \right\}.$$

a variável aleatória $D(y; \hat{\pi})$ é distribuída aproximadamente como χ_{n-p}^2 , onde p é o número de parâmetros ajustados segundo o modelo em investigação.

4 RESULTADOS E DISCUSSÃO

Entre 452 prontuários revisados foram selecionados 62 pacientes com Filariose Bancroftiana, dos quais 28 apresentavam síndrome de fistulização. O tempo gasto na entrevista social foi em média 60 minutos. A variável *gordura* foi dicotomizada, visto que não obtivemos bons resultados utilizando-a como variável contínua. O ponto de corte para análise em relação a ingestão de gordura foi considerado o de uma dieta normal de até 900g de lipídeos/mês. Para se estimar esse valor de gordura total consumida, usou-se o cut of de 400g de gordura adicionada ao alimento em forma de óleos, manteigas, margarinas, banhas, cujo valores foram obtidos nas entrevistas. A esse valor, foi calculado uma ingestão de cerca de 500g/mês proveniente da gordura “intrínseca” dos alimentos, consumida pela nossa população resultando em um consumo de 900g/mês. O consumo da proteína em nosso meio está muito associada ao lipídio “invisível” contido em salsichas, lingüiças, mortadelas, carnes gordas e ovos, além do uso do côco na preparação de muitos alimentos apreciados pelo nordestino. Inicialmente um total de 5 variáveis explicativas de natureza qualitativa e quantitativa foram utilizadas, sendo expressas por:

- SINDROME (1: paciente apresenta síndrome de fistulização, 0: se não apresenta)
- SEXO (0: masculino, 1: feminino);
- IDADE, idade em anos;
- PESO, peso em kg;
- GORDURA (0: até 400gr, 1: acima de 400gr);
- RENDA, em reais.

Uma vez definido o conjunto de covariáveis (ou fatores) a ser incluído num modelo binomial, resta saber qual a melhor maneira de encontrar um modelo reduzido que inclua apenas as covariáveis mais importantes para explicar a probabilidade de sucesso $\pi(x)$.

Inicialmente, ajustamos um modelo com todas as variáveis incluídas e, a partir da análise de uma seqüência de modelos encaixados (ver Seção 2.4.3), podemos medir a importância de cada uma delas no modelo.

A respeito da função de ligação, utilizamos a ligação *logística* devido algumas características citadas na Seção 3.2.4.

Temos na Tabela 4.1 os resultados do ajustamento do modelo com todas as variáveis incluídas.

Efeito	Estimativa	Desvio Padrão
Constante	2,456565693	2,503097282
Sexo	2,226244024	0,995971319
Idade	0,005182381	0,026708075
Peso	-0,076146047	0,034355220
Gordura	1,870413306	1,009857508
Renda	0,002034619	0,001549223

Tabela 4.1: *Estimativas dos parâmetros referentes ao modelo logístico com todas as variáveis incluídas.*

A Tabela 4.2 apresenta a análise do desvio, de onde se verifica que as variáveis *idade* e *renda* apresentam desvio residual inferior a $\chi^2_{1,0.05} = 3,841$, sendo assim, excluídas do modelo. Muitas vezes variáveis consideradas biologicamente importantes não devem ser deixadas de lado pela sua falta de significância estatística. Assim, a seleção de um modelo logístico deve ser um processo conjugado de seleção estatística de modelos e bom senso.

Termos	g.l.	Desvio Residual	g.l.	Desvio Residual
Nulo			61	85,36870
Sexo	1	7,587195	60	77,78150
Idade	1	1,291093	59	76,49041
Peso	1	6,159515	58	70,33089
Gordura	1	4,169421	57	66,16147
Renda	1	1,948814	56	64,21266

Tabela 4.2: *Análise do desvio*

Após o ajuste citado anteriormente, procedemos então a um novo processo de estimação que fornece como modelo final

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = 2,8388 + 2,0473\text{sexo} - 0,0739\text{peso} + 1,9388\text{gordura} \quad (4.1)$$

com os resultados apresentados na Tabela 4.3. Como podemos observar, há indícios de que a chance de ter síndrome de fistulização seja maior para o sexo feminino do que para o sexo masculino, já que a estimativa do parâmetro relacionada a esse fato é positiva. Nota-se também que a chance de ter síndrome fistulizante diminui com o peso e há indicações de que a chance de ter síndrome de fistulização aumenta significativamente com a quantidade de consumo de gordura ingerida na dieta. Por outro lado, na Índia estudos comprovam que a restrição de gorduras pode significativamente diminuir a lipidúria. Lipidúria severa pode ocorrer mesmo com 25g de gordura por dia (Singh LK, Datta B, Dwivedi US, Singh PB. Dietary fats and chyluria. Indian J Urol 2005;21:50-54)

Efeito	Estimativa	Desvio Padrão
Constante	2,8387797	2,35693336
Sexo	2,0472986	0,96603489
Peso	-0,0738899	0,03322744
Gordura	1,9388198	0,97266118

Tabela 4.3: Estimativas dos parâmetros associados ao modelo logístico final.

Note-se que o desvio residual do modelo (66,53019), apresentado na Tabela 4.4, é inferior ao valor crítico $\chi^2_{58,0.05} = 76,7778$, o que nos leva a aceitá-lo em princípio. Além disso, para todas variáveis explicativas, seus respectivos desvios residuais apresentam-se superiores a $\chi^2_{1,0.05} = 3,841$, sinalizando que as mesmas são importantes para o modelo. O número reduzido de iterações pelo Método Escore de Fisher (4 iterações), necessárias para convergência das estimativas dos parâmetros, é outro sinal positivo para o modelo.

Termos	g.l.	Desvio Residual	g.l.	Desvio Residual
Nulo			61	85,36870
Sexo	1	7,587195	60	77,78150
Peso	1	6,127413	59	71,65409
Gordura	1	5,123893	58	66,53019

Tabela 4.4: Análise do desvio para o modelo logístico final.

Alguns gráficos de diagnóstico são apresentados nas Figuras 4.1 e 4.2. No gráfico dos resíduos r_{Di} , Figura 4.1, a maioria dos pontos cai dentro do intervalo $[-2, 2]$, com apenas uma observação, #6, fora do intervalo, porém muito próxima ao limite. O paciente #6 é do sexo masculino, tem síndrome de fistulização, pesa 77kg e consome até 400gr de gordura mensal. Pelos resultados das estimativas seria mais provável esperar de um paciente com esse perfil que não apresentasse síndrome de fistulização. Na Figura 4.2 temos o gráfico normal de probabilidades para o resíduo r_{Di} , e não notamos nenhum indício de que a distribuição utilizada seja inadequada.

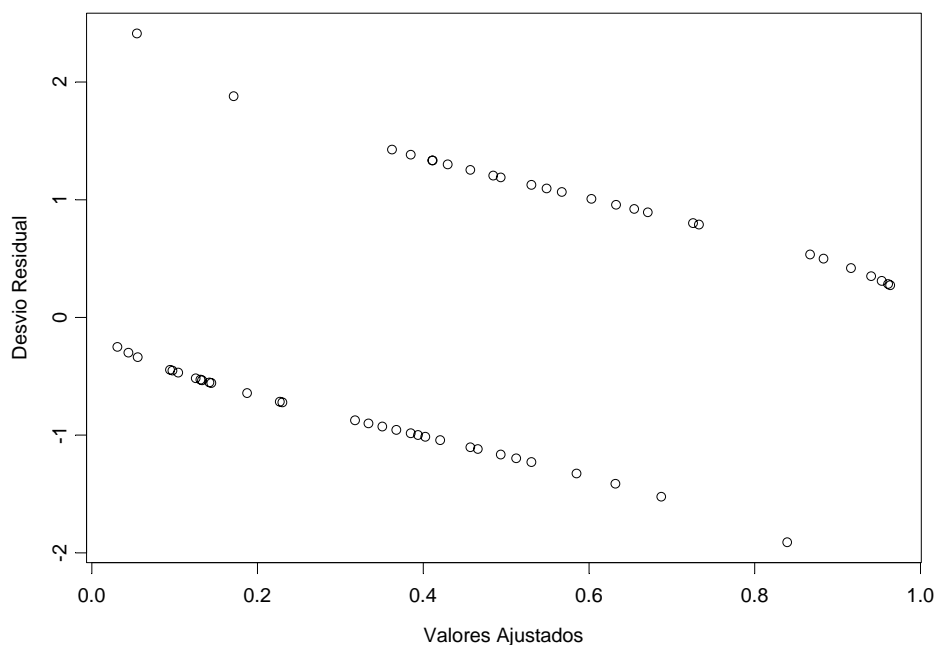


Figura 4.1: Representação gráfica dos resíduos versus valores ajustados.

Como o interesse principal é encontrar alguma relação entre os hábitos alimentares dos pacientes portadores de síndromes de fistulizantes e a quantidade de consumo de gordura ingerida na dieta, formamos a razão de chances envolvendo essa covariável. A razão de chances da síndrome de fistulização entre um paciente que consome gordura acima de 400gr e um paciente que consome gordura até 400gr, que denotaremos por ψ_{GO} , supondo que os pacientes tenham o mesmo sexo e peso, é estimada por

$$\hat{\psi}_{GO} = \exp\{1,9388\} = 6,95.$$

Logo, podemos concluir que a chance da síndrome de fistulização é aproximadamente 7 vezes maior para pacientes que consomem gordura acima de 400gr do que para pacientes que consomem gordura até 400gr, quando ambos tenham o mesmo sexo e peso.

Sabendo de antemão que qualquer mudança de hábito é um processo de aprendizagem e que mudar hábitos alimentares envolve muito mais do que substituir gêneros ou forma de fazê-los, consideramos também importante identificar o responsável pela escolha e confecção das refeições e sua posição dentro do grupo familiar como um dado imprescindível ao planejamento interdisciplinar para a educação alimentar.

Analogamente, seja ψ_{se} a razão de chances da síndrome de fistulização entre um paciente do sexo feminino e um paciente do sexo masculino. Supondo que os pacientes são semelhantes nas demais covariáveis (*peso e gordura*) esse parâmetro é estimado por

$$\hat{\psi}_{se} = \exp\{2,0473\} = 7,75.$$

Portanto, dessa expressão podemos deduzir que a chance de síndrome de fistulização é aproximadamente 7,75 vezes maior para pacientes do sexo feminino do que para pacientes do sexo masculino.

Já com relação a variável *peso*, não faz sentido calcular a chance de um indivíduo que tem um determinado peso p^* , ter síndrome de fistulização, em relação a um outro indivíduo que tenha peso p . Visto que, pacientes portadores de síndromes fistulizantes têm uma importante perda de peso. Logo, o peso não é um fator para desenvolvimento da síndrome e sim, uma consequência da mesma.

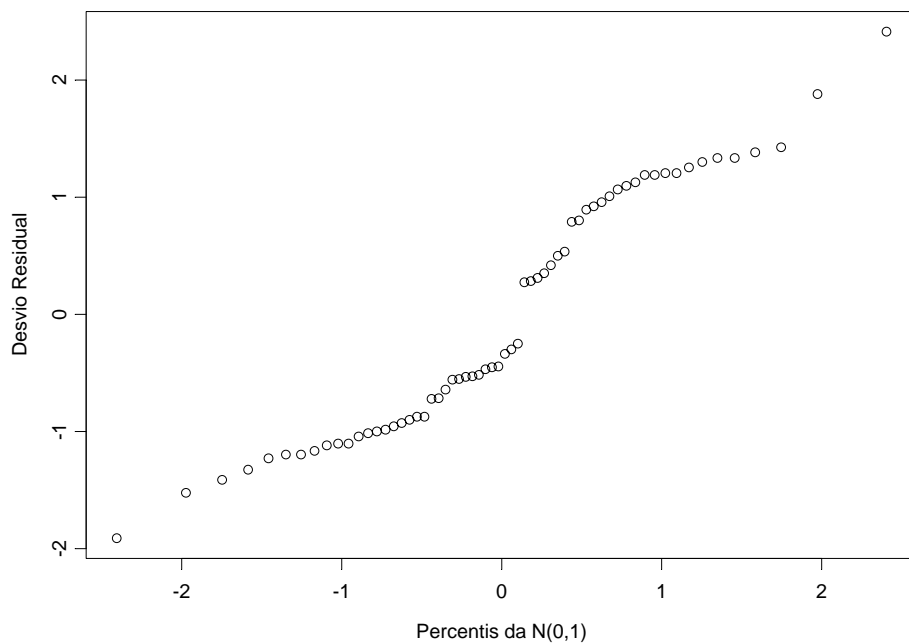


Figura 4.2: Representação gráfica dos resíduos versus percentis da $N(0, 1)$

Na Tabela 4.5 podemos observar a probabilidade de um paciente com um determinado conjunto de valores para as covariáveis *sexo*, *peso* e *gordura* estar com síndrome de fistulização.

Sexo	Peso	Gordura	$\hat{\pi}$
feminino	67,0	> 400gr	0,867
masculino	72,3	> 400gr	0,362
masculino	51,0	> 400gr	0,733
feminino	58,8	até 400gr	0,632
masculino	85,0	até 400gr	0,031

Tabela 4.5: Probabilidade para algumas configurações dadas.

5 CONCLUSÕES

As principais conclusões da análise apresentada na Seção anterior são as seguintes:

1. O modelo que melhor ajustou os dados e portanto explica melhor a probabilidade de de um paciente com filariose bancroftiana ter síndrome de fistulização é dado por:

$$\log \left(\frac{\pi_i}{1 - \pi_i} \right) = 2,8388 + 2,0473sexo - 0,0739peso + 1,9388gordura$$

2. O modelo logístico permitiu estabelecer a relação existente entre pacientes portadores de síndromes fistulizantes e as variáveis, sexo, peso e a quantidade de consumo de gordura ingerida na dieta.
3. Pacientes portadores de filariose bancroftiana do sexo feminino, tem aproximadamente 7,75 vezes mais chance de desenvolver a síndrome de fistulização do que pacientes do sexo masculino. Na nossa cultura cabe mais a mulher a responsabilidade de comprar o alimento, e praticamente são as mulheres que preparam o alimento, estando assim as mesmas mais expostas a gordura ao provarem o mesmo durante o cozimento. Por outro lado, o sistema linfático feminino, por razões fisiológicas, como a gravidez, ou por maior severidade de varizes venosas, é mais fragilizado, aumentando assim os riscos de danos linfáticos, podendo influenciar e justificar esse maior risco encontrado nesse estudo preliminar.
4. A chance de ter a síndrome de fistulização é aproximadamente 7 vezes maior para pacientes que consomem gordura acima de 400gr do que para pacientes que consomem gordura até 400gr. Logo, a mudança de hábitos alimentares, ou melhor, a quantidade de consumo de gordura ingerida na dieta é um fator fundamental para o não desenvolvimento da síndrome fistulizante. Como a qualidade da gordura consumida pela população acometida de filariose está composta na sua grande

maioria pelas gorduras malélicas (chamadas de gordura trans), a dieta hipolípídica com a devida orientação médica-nutricional de acordo com a realidade social do paciente nessa população deverá trazer a médio e longo prazos conseqüências benéficas relacionadas a possibilidade de diminuir os riscos para as doenças cardiovasculares, tão comuns em nossa população. Esse trabalho abre assim, novos horizontes para as pesquisas futuras em outras áreas do conhecimento, além da possibilidade da profilaxia das síndromes de fistulização.

5. O modelo logístico final poderá propiciar como diagnóstico inicial, a probabilidade de um paciente portador de filariose bancroftiana, estar com síndrome de fistulização.

6 REFERÊNCIAS

- 1 Aitkin, M., Anderson, D., Francis, B. e Hinde, J. (1989). *Statistical Modelling in GLIM*. Clarendon Press, Oxford, U.K.
- 2 Amaral, F., Dreyer, G., Figueredo-Silva, J., Norões, J., Cavalcanti, A., Samico, S.F., Santos, A., Coutinho, A., (1994). Live Adult Worms Detected by Ultrasonography in Human Bancroftian Filariasis. *American Journal of Tropical Medicine and Hygiene*, **50**: 753-757
- 3 Anscombe, F.J. (1953). Contribution to the discussion of H. Hotelling's paper. *J.R. Statist. Soc. B*, **15**, 229-230.
- 4 Aranda-Ordaz, F. (1981). On the families of transformations to additivity for binary response data. *Biometrika*, **68**, 357-363.
- 5 Barndorff-Nielsen, O.E. (1978). *Information and exponential families in statistical theory*. Wiley, Chichester.
- 6 Bernoulli, J. (1713). *Ars conjectandi*. Thurnisius, Basilea.
- 7 Bliss, C.I. (1935). The calculator of the dosage-mortality curve. *Ann. Appl. Biol.*, **22**, 134-167.
- 8 Breslow, N.E. e Clayton, D.G. (1993), Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9-25.
- 9 Collet, D. (1994). *Modelling binary data*. Chapman and Hall, London.
- 10 Cordeiro, G.M. (1986). *Modelos lineares generalizados*. VII SINAPE, UNICAMP.

- 11 Cordeiro, G.M. e Paula, G.A. (1992). Estimation, large-samples parametric tests and diagnostics for non-exponential family nonlinear models. *Communications in Statistics, Simulation and Computation*, **21**, 149-172.
- 12 Dreyer, G., Medeiros, Z., Netto, M.J., Leal, N.C., De Castro, L.G., Piessens, W.F., 1999. Acute attacks in the extremities of persons living in an area endemic for bancroftian filariasis: Differentiation of Two Syndromes. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 93: 413-417.
- 13 Dreyer, G., Norões, J., Figueredo-Silva, J., Piessens, W.F., 2000. Pathogenesis of Lymphatic Disease in Bancroftian Filariasis: A Clinical Perspective. *Parasitology Today*, 16: 544-548.
- 14 Dreyer, G., Brandão, A.C., Amaral, F., Medeiros, Z., Addiss, D., 1996. Detection by Ultrasound of Living Adult *Wuchereria bancrofti* in the Female Breast. *Mem. Inst. Oswaldo Cruz*, 91: 95-96.
- 15 Figueredo-Silva, J., Norões, J., Cedenho, A., Dreyer, G., 2002. Histopathology of Bancroftian Filariasis Revisited: the Role of the Adult Worm in the Lymphatic Vessel Disease. *Annals of Tropical Medicine and Parasitology*, 96: 531-541.
- 16 Fisher, R.A. (1925). *Statistical methods for research workers*. Oliver and Boyd, Edinburgh.
- 17 Fontes, G., Rocha, E.M.M., Brito, A.C. et al. 1998. Lymphatic filariasis in 32 Brazilian urban area (Maceió, Alagoas). *Mem Inst Oswaldo Cruz* 93:705-710.
- 18 Lane, P.W. e Nelder, J.A. (1982). Analysis of covariance and standardization as instances of prediction. *Biometrics*, **43**, 318-327.

- 19 Liang, K.Y. e Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.
- 20 McCullagh, P. e Nelder, J.A. (1983). *Generalized linear models*. Chapman and Hall, London.
- 21 Medeiros, Z., Dreyer, G., Andrade, L et al. 1992. *Wuchereria bancrofti* microfilarial density of autochthonous cases and natural *Culex* infectivity rates in Northeast Brazil. *J Trop Med Hyg* 95:214-217.
- 22 Medeiros, Z., Gomes, J., Beliz, F. et al. 1999. Screening of army soldiers for *Wuchereria bancrofti* infection in metropolitan Recife region, Brazil: implications for epidemiologic surveillance. *Trop Med Int Hlth* 4:499-505.
- 23 Michael, E., Bundy, D.A.P., Grenfell, B.T., 1996. Re-Assessing the Global Prevalence and Distribution of Lymphatic Filariasis. *Parasitology*, 112: 409-428.
- 24 Nelder, J.A. e Wedderburn, R.W.M (1972). *Generalized linear models*. *J. R. Statist. Soc. A*, **135**, 370-384.
- 25 Norões, J., Addiss, D., Cedenho, A., Figueredo-Silva, J., Lima, G., Dreyer, G., 2003. Pathogenesis of Filarial Hydrocele: Risk Associated with Intrascrotal Nodules Caused by Death of Adult *Wuchereria bancrofti*. *Trans. Roy. Soc. Trop. Med. Hyg.*, 97: 561-566.
- 26 Pierce, D.A. e Schafer, D.W. (1986). Residuals in generalized linear models. *Journal of the American Statistical Association*, **81**, 977-986.
- 27 Pregibon, D. (1979). *Data analytic methods for generalized linear models*. Unpublished Ph.D. thesis, Dept. of Statistics, University of Toronto.

- 28 Singh LK, Datta B, Dwivedi US, Singh PB. Dietary fats and chyluria. *Indian J Urol* 2005;21:50-54
- 29 Stukel, T.A. (1988). Generalized logistic models. *Journal of the American Statistical Association*, **83**, 426-431.
- 30 WHO - World Health Organization 1994. Strategies for control of lymphatic filariasis infection and disease: report of a WHO/CTD/TDR consultative meeting held at the Universiti Sains Malaysia. Penang, Malaysia, 22-24.
- 31 Yates, F. (1955). The use of transformations and maximum likelihood in the analysis of quantal experiments involving two treatments. *Biometrika*, **42**, 382-403.