

UM MODELO LOGÍSTICO-LINEAR APLICADO À FILARIOSE BANCROFTIANA

Fábio Cavalcanti PEREIRA¹
Gauss Moutinho CORDEIRO¹
Gerusa DREYER²
Denise MATTOS²

- RESUMO: O modelo linear generalizado (MLG) é definido por uma distribuição de probabilidade, membro da família exponencial de distribuições para a variável resposta, um conjunto de variáveis independentes descrevendo a estrutura linear do modelo e uma função de ligação relacionando a média da variável resposta e a estrutura linear. O objetivo do artigo é estimar a probabilidade de desenvolvimento das síndromes fistulizantes em pacientes portadores de filariose bancroftiana, uma doença que afeta mais de 100 milhões de pessoas no mundo, em função de algumas variáveis explicativas, principalmente, a quantidade de gordura na dieta. Para a análise foi usado o modelo logístico-linear ajustado através do software S-PLUS aos dados da doença. Conclui-se que a probabilidade de pacientes terem a síndrome de fistulização é, aproximadamente, 7 vezes maior quando consomem gordura acima de 400gr/mês, quando comparados com pacientes que consomem gorduras abaixo desse valor.
- PALAVRAS-CHAVE: Filariose bancroftiana; máxima verossimilhança; modelos lineares generalizados; modelo logístico; síndromes fistulizantes.

1 Introdução

O modelo linear generalizado (MLG), uma referência em português é o livro de Cordeiro (1986), é definido por uma distribuição de probabilidade para a variável resposta pertencente à família exponencial de distribuições, um conjunto

¹Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco – UFRPE, CEP: 52171-900, Recife, PE, Brasil. E-mail: *fabio_ufrpe@hotmail.com / gauss@deinfo.ufrpe.br*

²Núcleo de Ensino Pesquisa e Assistência em Filariose, Hospital das Clínicas, Universidade Federal de Pernambuco – UFPE, CEP: 50670-910, Recife, PE, Brasil. E-mail: *dreyer-g@uol.com.br / mattos.denise@gmail.com*

de variáveis independentes descrevendo a estrutura linear do modelo e uma função de ligação relacionando a média da variável resposta e a estrutura linear. Um caso particular importante destes modelos é o modelo logístico-linear em que a variável resposta é binária e a função de ligação é a logit.

A filariose bancroftiana, também conhecida por elefantíase, é uma doença exclusiva do homem causada por um parasito intravascular conhecido como *Wuchereria bancrofti*. É transmitida por um mosquito conhecido no Brasil como muriçoca e carapanã. Os vermes adultos vivem nos vasos linfáticos e linfonodos e, após o acasalamento, as fêmeas produzem um grande número de microfírias (Mf), que deixam os vasos linfáticos em direção à circulação sanguínea. Quando o mosquito se alimenta no ser humano portador de microfírias, as mesmas são sugadas e se transformam no que chamamos de larvas infectantes após um período médio de 14 dias. Por ocasião da próxima picada, a fêmea deposita as larvas na pele do indivíduo sadio. Essas larvas penetram pela picada do mosquito, migram até o sistema linfático, reiniciando, dessa forma, o ciclo. Assim, não ocorre transmissão de pessoa para pessoa, como uma parte da comunidade acredita. A bancroftose afeta, pelo menos, 100 milhões de pessoas distribuídas em 83 países dos diferentes continentes (WHO, 2003). Em nosso país, são considerados focos de transmissão ativa o Grande Recife, em Pernambuco (Medeiros et al., 1992) e a cidade de Maceió, em Alagoas (Fontes et al., 1998). Belém do Pará, que na década de 1950 era a área de maior prevalência, hoje é considerada uma área sem transmissão.

A infecção está geralmente limitada às populações de baixa renda, que vivem em condições sócio-econômicas extremamente precárias. Durrheim et al. (2004), mostram, de forma inequívoca, que a associação entre filariose bancroftiana e pobreza é um fato, quando 94% dos países com os mais baixos índices de desenvolvimento humano (IDH) são endêmicos. O substrato anátomo-patológico da infecção bancroftiana é a dilatação dos vasos linfáticos (linfangiectasia) causada pelos vermes vivos. Essa dilatação pode gerar diferentes conseqüências 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 de pele, cuja conseqüência é a forma mais desfigurante da doença, conhecida como elefantíase (Dreyer et al., 1999). Por outro lado, quando o linfático afetado está localizado nas partes mais internas do organismo e a bactéria não faz parte do processo, pode ocorrer a ruptura dos vasos linfáticos. Isso provoca o extravasando da linfa produzindo as síndromes de fistulização mais conhecidas como quilocele (linfa no interior da bolsa escrotal), quilúria (linfa na urina) e linfoescroto (perda de linfa através da pele) (Dreyer et al., 2000).

As síndromes fistulizantes, além de causarem debilitação física importante, produzem também estados de angústias e situações estigmatizantes na população afetada (Dreyer et al., 2005). Após uma refeição gordurosa, a pressão no interior dos vasos linfáticos fica extremamente elevada, provocando a perda de linfa rica em gordura. O tratamento constitui, essencialmente, em uma re-educação alimentar

(Dreyer et al., 2007). Essa mudança de hábitos alimentares estaria ligada ao consumo de uma dieta pobre em gorduras. Para os indivíduos vivendo em áreas endêmicas, seria importante identificar o(s) fator(es) de risco no sentido de evitar ou postergar as síndromes de fistulização. O objetivo do estudo foi verificar se existe ou não a associação entre a presença da síndrome fistulizante e a quantidade de consumo de gordura ingerida na dieta. Para encontrar tal relação foram ajustados modelos logístico-lineares. O software utilizado foi o S-PLUS que é um sistema de manipulação de dados, análise estatística e apresentação gráfica.

2 Modelos lineares generalizados e o modelo binomial

2.1 Modelo binomial

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

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

em que π_i é a probabilidade de sucesso. A distribuição binomial surge naturalmente em um grande número de contextos, sendo as observações contagens não-negativas limitadas por um valor fixo (o número de ensaios). Existem duas maneiras de deduzi-la. Sejam Y_1 e Y_2 variáveis aleatórias independentes com distribuição de Poisson com médias μ_1 e μ_2 , respectivamente. Sabe-se 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

$$P(Y_1 = y | Y_1 + Y_2 = m) = \binom{m}{y} \pi^y (1 - \pi)^{m-y}, \quad y = 0, 1, \dots, m, \quad (2)$$

onde $\pi = \frac{\mu_1}{(\mu_1 + \mu_2)}$. A notação $Y \sim B(m, \pi)$ denota que Y tem distribuição binomial, expressa em (2), com índice m e parâmetro π . A outra maneira vem da distribuição de Bernoulli, expressa em (1) quando $m = 1$, sendo, assim, um caso particular da distribuição binomial. A distribuição binomial é, então, uma soma estocástica $S_m = \sum_{i=1}^m Y_i$ de m variáveis aleatórias Y_1, \dots, Y_m de Bernoulli independentes e identicamente distribuídas.

2.2 Função geratriz de cumulantes

A função geratriz de momentos de (1) é

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

e, portanto, a função geratriz de cumulantes da distribuição de Bernoulli iguala

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

A função geratriz de cumulantes da binomial pode ser facilmente obtida a partir da soma de funções cumulantes de variáveis aleatórias de Bernoulli independentes. Portanto, a função geratriz de cumulantes da soma estocástica $S_m = Y_1 + \dots + Y_m$ é

$$K_{S_m}(t) = m \log\{1 - \pi + \pi exp(t)\}. \quad (3)$$

Expandindo (3) em série de Taylor e avaliando em $t = 0$, obtém-se os quatro primeiros cumulantes da distribuição binomial $\kappa_1 = m\pi$, $\kappa_2 = m\pi(1 - \pi)$, $\kappa_3 = m\pi(1 - \pi)(1 - 2\pi)$ e $\kappa_4 = m\pi(1 - \pi)\{1 - 6\pi(1 - \pi)\}$.

2.3 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) deve-se considerar a dependência entre π e (x_1, \dots, x_p) através da estrutura linear

$$\eta = \sum_{j=1}^p \beta_j x_j. \quad (4)$$

Entretanto, como $-\infty < \eta < +\infty$, expressar π através de uma função linear de η seria incorreto, pois π não ficaria restrito ao intervalo $(0, 1)$. Uma maneira simples e eficaz de solucionar este problema é usar uma função de ligação $g(\pi)$ que conduza o intervalo unitário à reta real, de modo que

$$g(\pi_i) = \eta_i = \sum_{j=1}^p x_{ij} \beta_j, \quad i = 1, \dots, n. \quad (5)$$

As três funções de ligação usualmente empregadas no modelo binomial que garantem a restrição do intervalo $(0, 1)$ imposta à probabilidade π são: logit ou função logística $g_1(\pi) = \log\left(\frac{\pi}{1 - \pi}\right)$; probit ou função inversa da distribuição acumulada da normal reduzida $g_2(\pi) = \Phi^{-1}(\pi)$; e complemento log-log $g_3(\pi) = \log\{-\log(1 - \pi)\}$. Estas três funções são contínuas e crescentes no intervalo $(0, 1)$. A função logística possui algumas características que a tornam preferida em relação às outras duas, tais como: pode ser interpretada como o logaritmo da razão de

chances, apresenta propriedades mais simples e é mais conveniente para análise de dados coletados de forma retrospectiva. A ligação logística é bastante empregada em estudos toxicológicos e epidemiológicos. Entretanto, isto não quer dizer que as outras transformações não sejam utilizadas na prática. Bliss (1935) iniciou a modelagem de proporções, utilizando um modelo binomial com ligação probit. A ligação complemento log-log é recomendada por Collet (1994) quando a distribuição das proporções é assimétrica.

Para se compreender melhor o ajuste obtido é necessário utilizar a relação entre π e o preditor linear $\eta = X\beta$. A função de ligação logit $\log\left(\frac{\pi}{1-\pi}\right) = \eta$ pode ser expressa em termos do preditor linear como $\pi = \frac{e^\eta}{1+e^\eta}$.

Assim, 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, essa probabilidade tende para um. Similarmente, pode-se calcular a relação entre π e η para as outras duas funções de ligação: $\pi = g_2^{-1}(\eta) = \Phi(\eta)$ e $\pi = g_3^{-1}(\eta) = 1 - \exp\{\exp(-\eta)\}$.

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

$$T_\lambda(\pi) = \frac{2\pi^\lambda - (1-\pi)^\lambda}{\lambda\pi^\lambda + (1-\pi)^\lambda}, \quad (6)$$

sendo que λ representa o parâmetro da transformação. Duas características importantes de (6) são $T_\lambda(\pi) = -T_\lambda(1-\pi)$ e $T_\lambda(\pi) = -T_{-\lambda}(\pi)$, ou seja, T_λ trata o sucesso e fracasso de forma simétrica. A família $T_\lambda(\pi)$ é denominada de simétrica. Para $\lambda = 0$, (6) se reduz à transformação logística, enquanto que para $\lambda = 1$ obtém-se a transformação linear. Quando for apropriado tratar sucesso e fracasso de forma assimétrica, uma segunda família de transformações proposta por Yates (1955) é definida por

$$W_\lambda(\pi) = \frac{\{(1-\pi)^{-\lambda} - 1\}}{\lambda}. \quad (7)$$

A expressão (7) iguala a transformação logística quando $\lambda = 1$ e a complemento log-log quando $\lambda = 0$. Para maiores detalhes sobre estas famílias de ligações, vide Aranda-Ordaz (1981). Além disso, torna-se importante ressaltar que o trabalho de Stukel (1988) generaliza a família apresentada por Aranda-Ordaz (1981). Por motivos de simplicidade de interpretação a ligação logística será adotada nesse artigo. Entretanto, o leitor poderá encontrar estudos comparativos entre as funções de ligação logística, probit e complemento log-log no livro de McCullagh e Nelder (1989).

2.4 Função de verossimilhança

Considera-se que os dados y_1, \dots, y_n são valores observados de variáveis aleatórias independentes Y_1, \dots, Y_n com distribuição binomial de índice m_i e

parâmetro π_i , respectivamente. De (2), tem-se que a log-verossimilhança de π dado y pode ser escrita, sem o termo combinatório que não envolve o parâmetro π , como

$$l(\pi; y) = \sum_{i=1}^n \left[y_i \log \left(\frac{\pi_i}{1 - \pi_i} \right) + m_i \log(1 - \pi_i) \right]. \quad (8)$$

Escolhendo para o modelo binomial a função de ligação logística

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

a log-verossimilhança (8) em função dos parâmetros β iguala

$$l(\beta; y) = \sum_{i=1}^n \sum_{j=1}^p y_i x_{ij} \beta_j - \sum_{i=1}^n m_i \log \left\{ 1 + \exp \left(\sum_{j=1}^p x_{ij} \beta_j \right) \right\}. \quad (9)$$

Um ponto importante que deve ser ressaltado é que a estatística $X^T y$, que aparece na log-verossimilhança (9), é suficiente para β , pois a ligação logística é, também, a ligação canônica no modelo binomial.

2.5 Estimação dos parâmetros

A estimação dos parâmetros usa o método escore de Fisher para parâmetros. A função escore e a matriz de informação são

$$U(\beta) = X^T(y - \mu) \text{ e } K = X^T W X,$$

em que $W = \text{diag} \{m_i \pi_i (1 - \pi_i)\}$. O algoritmo iterativo resultante para obter a estimativa de máxima verossimilhança (EMV) de β é dado por

$$\beta^{(m+1)} = \beta^{(m)} + K^{(m)^{-1}} U(\beta^{(m)}). \quad (10)$$

É importante salientar que no processo iterativo (10) as observações com maior variância $V(\pi_i) = m_i \pi_i (1 - \pi_i)$, apresentam maior peso w_i para o cálculo da EMV do vetor β .

2.6 A função desvio

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

$$l(\hat{\pi}; y) = \sum_{i=1}^n \{y_i \log \hat{\pi}_i + (m_i - y_i) \log(1 - \hat{\pi}_i)\}, \quad (11)$$

em que $\hat{\pi}_i = \hat{\mu}_i/m_i$ e $\hat{\mu}_i = m_i\hat{\pi}_i$ é a média ajustada. No modelo saturado, a EMV de π_i é obtida de $\hat{\pi}_i = y_i/m_i$. Assim, a função desvio $D(y; \hat{\pi}) = 2l(\hat{\pi}; y) - 2l(\hat{\pi}; y)$ para o modelo binomial é expressa por

$$D(y; \hat{\pi}) = 2 \sum_{i=1}^n \left\{ y_i \log(y_i/\hat{\mu}_i) + (m_i - y_i) \log \left(\frac{m_i - y_i}{m_i - \hat{\mu}_i} \right) \right\}. \quad (12)$$

A estatística $D(y; \hat{\pi})$ é distribuída, aproximadamente, como χ_{n-p}^2 , sendo p o número de parâmetros ajustados segundo o modelo em investigação. Cordeiro (1995) estuda a adequação dessa aproximação em pequenas amostras.

2.7 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, considerando nos modelos a mesma função de ligação e distribuição, visando obter os efeitos de fatores, covariáveis e suas possíveis interações. Dois modelos são considerados encaixados quando os termos que formam o maior modelo M_{ps} incluem todos os termos que compõem M_{pr} mais outros termos que não estão em M_{pr} .

Considere-se uma seqüência de r modelos encaixados $M_{p1} \subset M_{p2} \subset \dots \subset M_{pr}$, com respectivas dimensões $p_1 < p_2 < \dots < p_r$, matrizes modelos $X_{p1}, X_{p2}, \dots, X_{pr}$ e desvios $D_{p1} > D_{p2} > \dots > D_{pr}$, todos os modelos com a mesma distribuição e a mesma função de ligação. As diferenças entre os desvios $D_{pi} - D_{pj}$, $p_i < p_j$, devem ser interpretadas como uma medida de variação dos dados, explicada pelos termos que estão em M_{pj} e não estão em M_{pi} . Se $D_{pi} - D_{pj} > \chi_{p_j - p_i, \alpha}^2$ consideramos que os termos que estão em M_{pj} e não estão em M_{pi} são significativos. Para maiores informações, vide Cordeiro (1986) e McCullagh e Nelder (1989). Deve-se ressaltar que as desigualdades entre os desvios não são válidas para a estatística de Pearson generalizada definida por $X^2 = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / \hat{V}_i$, em que \hat{V}_i é a função de variância estimada. Em suma, a comparação de modelos encaixados é realizada, exclusivamente, pela função desvio (12).

3 Metodologia e resultados

O estudo proposto é retrospectivo e 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, de acordo com os critérios de inclusão e exclusão listados abaixo. As informações foram extraídas de prontuários dos pacientes atendidos pelo médico assistente e pelo assistente social. Todos os pacientes assinaram o termo de consentimento geral permitindo o uso de seus dados, para fins científicos, desde que a sua identidade não fosse revelada. 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 e procedentes do Grande Recife;
- Ter o seu peso anotado;
- Ter entrevista para quantificação de ingestão de gordura quantificável constando o conteúdo das três principais refeições e de merendas, consumo mensal de óleo e de margarina, manteiga ou banha;
- Ter a entrevista da pessoa que prepara o alimento e/ou faz as compras da casa;
- Ter definido o diagnóstico clínico/laboratorial da forma clínica filarial.

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 ingesta lipídica;
- Ter outras doenças em que houvesse indicação de 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 processados como frituras, entre outros).

Entre 452 prontuários revisados foram selecionados 62 pacientes portadores de filariose bancroftiana - 34 portadores de formas não fistulizantes e 28 com síndrome de fistulização. O tempo gasto na entrevista social foi em média 60 minutos. A variável resposta *síndrome* indica se o paciente apresenta ou não síndrome de fistulização. Um total de 5 variáveis explicativas de natureza qualitativa e quantitativa foi coletado, a saber: *sexo* (0: masculino, 1: feminino); *idade*, idade em anos; *peso*, peso em kg; *renda*, em reais; e *gordura* (0: até 400gr/mês, 1: acima de 400gr/mês). A variável gordura foi dicotomizada, pois não se obteve bons resultados utilizando-a como variável contínua. As estatísticas descritivas das variáveis explicativas são apresentadas nas Tabelas 1 a 3.

Inicialmente, ajusta-se um modelo incluindo todas as variáveis explicativas adotando-se a função de ligação logística devido a algumas facilidades citadas na Seção 2.3.

Tabela 1 - Estatística descritiva da variável *sexo*

<i>Sexo</i>	%
M	82,30%
F	17,70%

Tabela 2 - Estatística descritiva da variável *gordura*

<i>Gordura</i>	%
até 400	17,70%
acima 400	82,30%

Tabela 3 - Estatísticas descritivas das variáveis: *idade, peso e renda*

Variáveis	Média	Mediana	Desvio Padrão
<i>Idade</i>	34,98	32,00	12,2267
<i>Peso</i>	68,08	67,00	11,2276
<i>Renda</i>	191,65	111,00	234,0023

A Tabela 4 apresenta as estimativas, erros padrão e a estatística t, em que se verifica que idade e renda não são estatisticamente significativas.

Tabela 4 - Estimativas dos parâmetros referentes ao modelo logístico com todas as variáveis incluídas

Efeito	Estimativa	Erro Padrão	Estatística t
<i>Constante</i>	2,456565693	2,503097282	0,9814104
<i>Sexo</i>	2,226244024	0,995971319	2,2352491
<i>Idade</i>	0,005182381	0,026708075	0,1940380
<i>Peso</i>	-0,076146047	0,034355220	-2,2164331
<i>Gordura</i>	1,870413306	1,009857508	1,8521557
<i>Renda</i>	0,002034619	0,001549223	1,3133154

A partir da análise de uma seqüência de modelos encaixados como apresentado na Seção 2.7, pode-se medir a importância de cada uma das variáveis explicativas do modelo. A Tabela 5 apresenta a análise do desvio. Enfatizamos que as diferenças entre os desvios residuais têm, aproximadamente, distribuição qui-quadrado, pois representam razões de verossimilhanças. Verifica-se, então, que as variáveis idade e renda apresentam diferenças residuais inferiores ao valor crítico $\chi^2_{1,0.05} = 3,841$ baseado na tabela qui-quadrado, sendo assim, excluídas do modelo. Muitas vezes, algumas 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.

Tabela 5 - Análise do desvio

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

Após o ajuste inicial, procede-se a um novo ajuste sem as variáveis explicativas *idade* e *renda*, conduzindo ao modelo estimado

$$\log\left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}\right) = 2,8388 + 2,0473 \textit{sexo} - 0,0739 \textit{peso} + 1,9388 \textit{gordura}. \quad (13)$$

Das estimativas dos parâmetros apresentadas na Tabela 6 pode-se observar que há indícios de que a chance de ter síndrome de fistulização seja maior para o sexo feminino do que para o sexo masculino, pois a estimativa do parâmetro correspondente é positiva e o nível sexo feminino é um. Nota-se, também, que a chance de ter síndrome fistulizante diminui com o peso e há um forte indicativo de que aumente, 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 25gr de gordura por dia (Singh et al., 2005).

Tabela 6 - Estimativas dos parâmetros associados ao modelo logístico final

Efeito	Estimativa	Erro Padrão	Estatística t
<i>Constante</i>	2,8387797	2,35693336	1,204438
<i>Sexo</i>	2,0472986	0,96603489	2,119280
<i>Peso</i>	-0,0738899	0,03322744	-2,223761
<i>Gordura</i>	1,9388198	0,97266118	1,993315

O desvio residual do modelo (66,53019), apresentado na Tabela 7, é inferior ao valor crítico $\chi_{58,0.05}^2 = 76,7778$, o que leva a aceitá-lo globalmente. Além disso, para todas variáveis explicativas, suas respectivas diferenças apresentam-se superiores ao valor crítico $\chi_{1,0.05}^2 = 3,841$, sinalizando que as mesmas são importantes para o modelo. O número reduzido de iterações necessárias para convergência das

estimativas dos parâmetros pelo método escore de Fisher (4 iterações) é outro indicativo de aceitação do modelo.

Adotam-se os resíduos definidos como as raízes quadradas sinalizadas do desvio, geralmente denotados por r_{Di} , pois esses resíduos têm, aproximadamente, a distribuição normal reduzida (Cordeiro, 1986; McCullagh e Nelder, 1989) independente da distribuição ser discreta ou contínua. A análise desses resíduos no contexto dos MLGs equivale - de forma aproximada -, aos resíduos do modelo normal linear e o leitor poderá consultar os dois livros anteriormente citados.

Tabela 7 - Análise do desvio para o modelo logístico final

Modelo	g.l	Diferença	g.l	Desvio Residual
<i>Constante</i>			61	85,36870
<i>Sexo</i>	1	7,587195	60	77,78150
<i>Peso</i>	1	6,127413	59	71,65409
<i>Gordura</i>	1	5,123893	58	66,53019

A Figura 1 apresenta o gráfico dos resíduos r_{Di} versus os valores ajustados. A maioria dos pontos fica dentro do intervalo $[-2, 2]$, com apenas uma observação, aquela #6, fora do intervalo, porém muito próxima do limite superior. O paciente #6 é do sexo masculino, tem síndrome de fistulização, pesa 77kg e consome até 400gr de gordura mensal. As estimativas indicam que seria mais provável esperar de um paciente com esse perfil que não apresentasse síndrome de fistulização. Por outro lado, os autores especulam que como não foi mensurado o seu dano linfático, pode-se aventar a possibilidade que pacientes com um dano linfático mais extenso, síndromes de fistulização ocorram mesmo em pacientes ingerindo menos de 400gr de gordura mensalmente.

A Figura 2 apresenta o gráfico dos desvios residuais ordenados versus os percentis da distribuição normal reduzida. O padrão esperado do gráfico próximo à primeira bissetriz mostra que não há indícios de que a distribuição binomial utilizada seja inadequada.

Pela compreensão da fisiopatologia da doença tornou-se interesse principal verificar se existe ou não alguma relação entre os hábitos alimentares dos pacientes portadores de síndromes de fistulizantes, em especial o consumo (quantidade) de gordura ingerida na dieta. Neste sentido, formamos a razão de chances envolvendo essa co-variável. A razão de chances da síndrome de fistulização entre um paciente que consome gordura acima de 400gr/mês e um paciente que consome gordura até 400gr/mês, 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, conclui-se que a probabilidade da síndrome de fistulização é, aproximadamente, 7 vezes maior para pacientes que consomem gordura acima de 400gr/mês do que para pacientes que consomem gordura até 400gr/mês, quando ambos têm o mesmo sexo e peso.

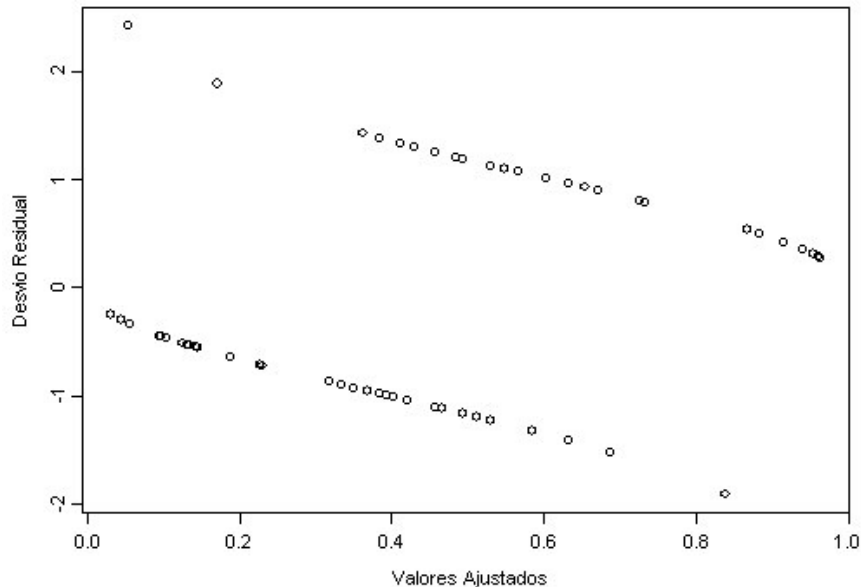


Figura 1 - Representação gráfica dos resíduos versus valores ajustados.

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, considera-se, 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 (Dreyer et al., 2007; Mattos e Dreyer, 2006).

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, pode-se 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.

Em relação à 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.

A Tabela 8 apresenta a probabilidade de um paciente - com um determinado conjunto de valores para as variáveis *sexo*, *peso* e *gordura* -, estar com síndrome de fistulização. Nesses termos, o médico com base nos valores dessas variáveis explicativas para qualquer paciente, poderá estimar a chance dele estar com síndrome de fistulização.

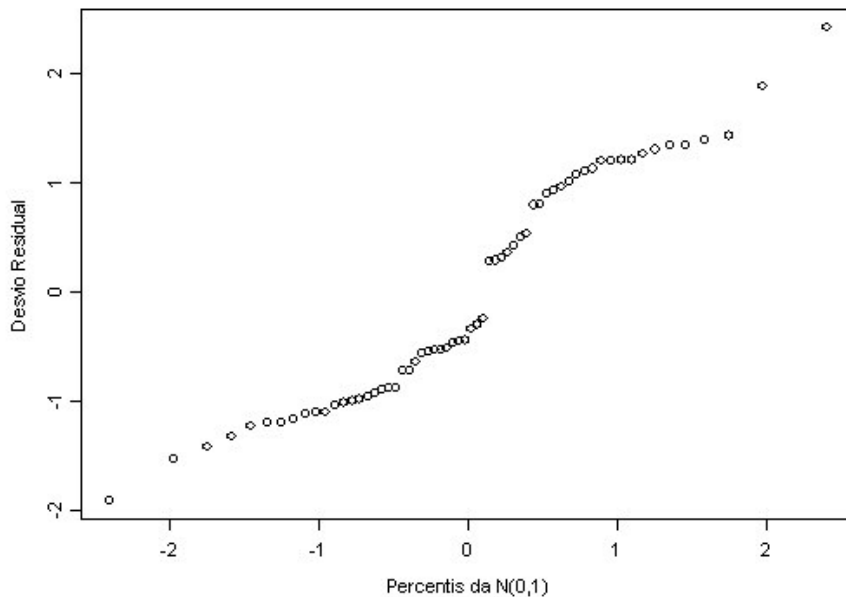


Figura 2 - Representação gráfica dos resíduos versus percentis da $N(0,1)$.

Tabela 8 - Probabilidades para algumas configurações dadas

Sexo	Peso	Gordura	$\hat{\pi}$
Feminino	67,0	acima 400	0,867
Masculino	72,3	acima 400	0,362
Masculino	51,0	acima 400	0,733
Feminino	58,8	até 400	0,632
Masculino	85,0	até 400	0,031

Conclusões

1. O modelo que melhor se ajusta aos dados e, portanto, explica melhor a probabilidade de um paciente com filariose bancroftiana ter síndrome de fistulização é dado pela equação de estimação (13);
2. O modelo logístico-linear permite estabelecer a relação existente entre pacientes portadores de síndromes fistulizantes e as variáveis explicativas, sexo, peso e a quantidade de consumo de gordura ingerida na dieta. Assim, pode-se verificar a sensibilidade da probabilidade de um paciente com filariose bancroftiana ter síndrome de fistulização em função das variáveis acima;
3. Pacientes portadores de filariose bancroftiana do sexo feminino têm,

aproximadamente, 7,75 vezes mais chance de desenvolver a síndrome de fistulização do que pacientes do sexo masculino. Na nossa cultura cabe mais à mulher a responsabilidade de comprar o alimento, e praticamente são as mulheres que preparam o alimento, estando assim as mesmas mais expostas à 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 cerca 7 vezes maior para pacientes que consomem gordura acima de 400gr/mês quando comparados com pacientes que consomem gordura abaixo desse valor. Logo, a mudança de hábitos alimentares, ou melhor, a quantidade de consumo de gordura ingerida na dieta parece ser um fator importante 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 de origem animal, a dieta hipolípídica com a devida orientação médico-nutricional, de acordo com a realidade social do paciente nessa população, deverá trazer a médio e longo prazos consequências benéficas relacionadas à possibilidade de diminuir os riscos para as doenças cardiovasculares, tão comuns em nossa população. Esse artigo abre assim novos horizontes para 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.

Agradecimentos

Agradecemos aos pacientes e aos seus familiares pela participação no estudo e a ONG Amaury Coutinho pelo suporte financeiro para o desenvolvimento do presente trabalho. Agradecemos, também, às várias sugestões de dois revisores que melhoraram a versão original do artigo.

PEREIRA, F. C.; CORDEIRO, G. M.; DREYER, G.; MATTOS, D. A logistic-linear model applied to Bancroftian filariasis. *Rev. Mat. Estat.*, São Paulo, v.25, n.4, p.7-22, 2007.

- **ABSTRACT:** The generalized linear model is defined by a probability distribution, member of the exponential family of distributions, for the response variable, a set of explanatory variables describing the linear structure of the model and a link function relating the mean response and the linear structure. Bancroftian filariasis is a major public health problem in the tropics with 100 millions of people infected. The aim of this investigation is to study the relationship between the risk of developing fistulation syndrome and several explanatory variables, such as, sex, age, weight, income and an amount of fat on the diet of patients assisted at the 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 logistic linear models fitted with the software S-PLUS. The study concludes that the probability of patients to have the fistulization syndrome is, approximately, 7 bigger when they consume fat above of 400gr/month, when compared with patients who consume fat below of this value.
- **KEYWORDS:** Bancroftian filariasis; fistulization syndromes; generalized linear model; logistic model; maximum likelihood.

Referências

- ARANDA-ORDAZ, F. On the families of transformations to additivity for binary response data. *Biometrika*, London, v.68, p.357-363, 1981.
- BLISS, C. I. The calculator of the dosage-mortality curve. *Ann. Appl. Biol.*, Wellesbourne, v.22, p.134-167, 1935.
- COLLET, D. *Modelling binary data*. London: Chapman and Hall, 1994.
- CORDEIRO, G. M. Modelos lineares generalizados. In: SINAPE, 7, 1986, Campinas. *Anais do Evento...* Campinas: ABE, 1986.
- CORDEIRO, G. M. Performance of a Bartlett-type modification for the deviance. *J. Stat. Comp. Simul.*, New York, v.51, p.383-403, 1995.
- DREYER, G. et al. Acute attacks in the extremities of persons living in an area endemic for bancroftian filariasis: differentiation of two syndromes. *Trans. R. Soc. Trop. Med. Hyg.*, London, v.93, p.413-417. 1999.
- DREYER, G. et al. Pathogenesis of lymphatic disease in bancroftian filariasis: a clinical perspective. *Parasitol. Today*, Cambridge, v.16, p.544-548, 2000.
- DREYER G.; MATTOS, D.; NORÕES, J. Filariose bancroftiana. In: COURA, J. R. (Ed.). *Dinâmica das doenças infecciosas e parasitárias*. Rio de Janeiro: Guanabara Koogan. 2005. cap.91, p.1087-1106.
- DREYER G.; MATTOS, D.; NORÕES, J. Quilúria. *Rev. Assoc. Med. Bras.*, São Paulo, v.53, p.460-464, 2007.
- DURRHEIM, D. N. et al. Lymphatic filariasis endemicity - an indicator of poverty? *Trop. Med. Int. Health*, Oxford, v.9, p.843-845, 2004.

- FONTES, G.; ROCHA, E. M. M.; BRITO, A. C. Lymphatic filariasis in 32 Brazilian urban area (Maceió, Alagoas). *Mem. Inst. Oswaldo Cruz*, Rio de Janeiro, v.93, p.705-710, 1998.
- MATTOS, D.; DREYER, G. A filariose bancroftiana: uma perspectiva psicossocial e política para os profissionais da saúde. *Rev. Panam. Infect.*, Santo Domingo, v.8, p.35-39, 2006.
- McCULLAGH, P.; NELDER, J. A. *Generalized linear models*. London: Chapman and Hall, 1989.
- MEDEIROS, Z.; DREYER, G.; ANDRADE, L. Wuchereria bancrofti microfilarial density of autochthonous cases and natural culex infectivity rates in northeast Brazil. *J. Trop. Med. Hyg.*, Oxford, v.95, p.214-217, 1992.
- SINGH, L. K. et al. Dietary fats and chyluria. *Indian J. Urol.*, Mumbai, v.21, p.50-54, 2005.
- STUKEL, T. A. Generalized logistic models. *J. Am. Stat. Assoc.*, New York, v.83, p.426-431, 1988.
- WHO - World Health Organization. *Global programme to eliminate LF: annual report on Lymphatic Filariasis*. 2003.
- YATES, F. The use of transformations and maximum likelihood in the analysis of quantal experiments involving two treatments. *Biometrika*, London, v.42, p.382-403, 1955.

Recebido em 10.10.2006.

Aprovado após revisão em 20.12.2007.