



**UNIVERSIDADE ESTADUAL DA PARAÍBA**  
**CENTRO DE CIÊNCIAS E TECNOLOGIA**  
**CAMPUS I - CAMPINA GRANDE - PB**  
**CURSO DE BACHALERADO EM ESTATÍSTICA**

Saullo Wesley Soares Da Silva

**Aplicação dos Modelos Lineares Generalizados  
com distribuição Binomial para dados Binários**

Campina Grande - PB

Março de 2014

Saullo Wesley Soares Da Silva

## **Aplicação dos Modelos Lineares Generalizados com distribuição Binomial para dados Binários**

Monografia apresentada ao Curso de Bacharelado em Estatística da Universidade Estadual da Paraíba, referente ao Trabalho de Conclusão de Curso (TCC), com exigência para a obtenção do título de graduação no Curso de Bacharelado em Estatística.

**Orientador:**  
**Tiago Almeida de Oliveira**

Campina Grande - PB

Março de 2014

É expressamente proibida a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da dissertação.

S586a Silva, Saullo Wesley Soares da.

Aplicação dos modelos lineares generalizados com distribuição binomial para dados binários [manuscrito] / Saulo Wesley Soares da Silva. - 2014.

47 p.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2014.

"Orientação: Prof. Dr. Tiago Almeida de Oliveira, Departamento de Estatística".

1. Superdispersão. 2. Família exponencial. 3. Função de ligação. I. Título.

21. ed. CDD 519.5

Saullo Wesley Soares Da Silva

## Aplicação dos Modelos Lineares Generalizados com distribuição Binomial para dados Binários

Monografia apresentada ao Curso de Bacharelado em Estatística da Universidade Estadual da Paraíba, referente ao Trabalho de Conclusão de Curso (TCC), com exigência para a obtenção do título de graduação no Curso de Bacharelado em Estatística.

Aprovada em: 18 / 03 / 2014

### Banca Examinadora



Prof. Tiago Almeida de Oliveira  
Orientador



Prof. Ricardo Alves de Olinda  
Universidade Estadual da Paraíba



Prof. Joherlan Campos de Freitas  
Universidade Estadual da Paraíba

## Dedicatória

Em especial aos meus pais, pelo apoio, pela dedicação, companheirismo e amizade, que tiveram comigo durante a concretização deste trabalho.

A minha namorada, que sempre me motivou para o término deste trabalho e que me faz uma pessoa mais feliz.

Aos meus familiares, exemplos de sucesso. E aos verdadeiros amigos.

## **Agradecimentos**

Agradeço a Deus por ter me capacitado em todos os momentos deste curso, por ter me dado forças, entendimento e sabedoria, sem Ele nada seria. A Ele a honra o poder e a glória.

Aos meus familiares, que compreenderam minha ausência dando força, amor, carinho e incentivos para que alcançasse mais esta vitória. Há vocês meu muito obrigado.

A minha namorada, por estar ao meu lado nos momentos de dificuldades e pelo apoio e confiança depositado.

Para que a concretização deste estudo se efetivasse: agradeço as inúmeras pessoas que foram incentivadoras neste processo seus ensinamentos serão a partir de agora essenciais em minha caminhada pessoal e profissional. Então, por estes extraordinários exemplos, expresso meus agradecimentos.

Ao orientador, Prof. Dr. Tiago Almeida de Oliveira por não medir esforços para conseguir material bibliográfico para a criação e desenvolvimento deste trabalho e total atenção, à você professor, meu muito obrigado.

Aos meus amigos, que me ajudaram a passar por bons e maus momentos, superar as dificuldades e me fizeram rir e descontraír nos momentos mais tensos. Obrigado!

Aos demais professores e funcionários do curso que, de uma forma ou outra, contribuíram para esta conquista, a todos meu agradecimento.

[...] não pergunte o que realmente sou; qual o meu verdadeiro eu; o que de essencial existe em mim. Pergunte como posso redescrever-me, de maneira a viver uma vida melhor ou mais bela. (RORTY, e COSTA, 1994, p. 21).

”O que é desconhecido é incerto, e toda incerteza deve ser quantificada em termos de probabilidade.”Thomas Bayes (1702 - 1761)

## Resumo

Os modelos lineares generalizados, foram desenvolvidos por Nelder e Wedderburn, que por sua vez estende-se o domínio da aplicabilidade dos modelos normais lineares e inclui modelos cuja variável resposta pertence à família exponencial de distribuições. Exemplos são os modelos log-lineares, modelos gama para dados contínuos, modelos logito e probito para dados de proporções, dentre outros. Nelder e Wedderburn desenvolveram um algoritmo para calcular as estimativas de máxima verossimilhança, que equivale a uma regressão do mínimos quadrados ponderados. Após realizar o ajuste de um modelo linear generalizado a determinado conjunto de observações, a variação residual poderá ser maior do que a variação que pode ser acomodada pelo modelo. Isso poderá indicar a ocorrência de superdispersão. Dentre várias das possíveis causas da superdispersão, duas delas podem ser a forma como o experimento foi conduzido e o processo de coleta dos dados. Quando ocorre a superdispersão, deve-se procurar variáveis explanatórias adicionais, ou incluir no modelo uma fonte adicional de variação que a acomode. Uma das maneiras de se acomodar a superdispersão é incluir um efeito aleatório no modelo. Neste trabalho, o objetivo é apresentar a estimação dos parâmetros de um modelo linear generalizado e sua implementação no software R.

**Palavras-chave:** Família exponencial; função de Ligação; anticoncepcional.



## Abstract

The generalized linear models were developed by Nelder and Wedderburn, which in turn extends the domain of applicability of linear normal models and includes models whose response variable belongs to the exponential family of distributions. Examples are the log - linear models, models range for continuous data, probit and logit models for data proportions, among others. After the setting of a generalized linear model to a given set of observations, the residual variation may be greater than the variation that can be accommodated by the model. This may indicate the occurrence of overdispersion. Among several possible causes of overdispersion, two of them may be the way the experiment was conducted and the data collection process. When overdispersion occurs, you should seek additional explanatory variables, or include in the model an additional source of variation that contemplate. In this work, the goal is to present the estimation of a linear model generalized applied to binary data using the binomial distribution and comparing functions for connection using the analysis software R.

**Keywords:** Exponential family; link function; contraceptive

# Lista de Figuras

- 1 Gráfico de resíduos para o modelo (3.1) com função de ligação logit p. 35
- 2 Gráfico de resíduos para o modelo (3.2) com função de ligação logit p. 36
- 3 Gráfico de resíduos para o modelo (3.1) sob função de ligação probit p. 38
- 4 Gráfico de resíduos para o modelo (3.2) sob função de ligação probit p. 39

# Lista de Tabelas

1	A Frequência do uso de contraceptivos entre mulheres casadas por idade, escolaridade e desejo de ter mais filhos (DMF), de Fiji Fertility Survey, 1975. . . . .	p. 32
2	Interação dos efeitos principais, Idade, Escolaridade e DMF, com ligação Logit. . . . .	p. 33
3	Interação dos efeitos principais, idade, escolaridade e DMF e suas interações, Idade e Educação, Idade e DMF, e Educação e DMF, com ligação Logit. . . . .	p. 34
4	Comparação dos modelos aninhados sob a função de ligação logit.	p. 35
5	Interação dos efeitos principais, Idade, Escolaridade e DMF, com ligação Probit. . . . .	p. 37
6	Interação dos efeitos principais, idade, escolaridade e DMF e suas interações, Idade e Educação, Idade e DMF, e Educação e DMF, com ligação Probit. . . . .	p. 38

# Lista de abreviaturas

**UEPB:** Universidade Estadual da Paraíba

**DMF:** Desejo de ter mais filho

**MLG:** Modelos Lineares Generalizados

**fdp:** função densidade de probabilidade

**fmp:** função massa de probabilidade

**fgm:** função geradora de momentos

**fgc:** função geradora de cumulantes

**aa:** amostra aleatória

**fda:** função de distribuição acumulada

**MV:** Máxima Verossimilhança

**EMV:** Estimação de Máxima Verossimilhança

# Sumário

Dedicatória . . . . .	
Agradecimentos . . . . .	
Resumo . . . . .	
Abstract . . . . .	
<b>1 Introdução</b>	<b>p. 1</b>
<b>2 Revisão de Literatura</b>	<b>p. 3</b>
2.1 Família Exponencial . . . . .	p. 3
2.2 Modelos Lineares Generalizados . . . . .	p. 7
2.3 Estimação . . . . .	p. 11
2.4 Distribuição Binomial . . . . .	p. 20
2.5 Transformação Logit . . . . .	p. 22
2.6 Transformação Probit . . . . .	p. 24
2.7 Teste de Hipóteses . . . . .	p. 26
2.7.1 Teste de Wald . . . . .	p. 26
2.8 Validação . . . . .	p. 27
2.8.1 Modelo Completo ou Saturado . . . . .	p. 27
2.9 Qualidade do Ajustamento . . . . .	p. 27

<b>3</b>	<b>Aplicação</b>	p. 30
3.1	Resultados . . . . .	p. 33
<b>4</b>	<b>Conclusão</b>	p. 40
	<b>Referências</b>	p. 41

# 1 Introdução

Em muitos estudos estatísticos, quer seja de natureza experimental ou observacional, com problemas em que o objetivo principal é o de estudar a relação entre variáveis, ou mais particularmente, analisar a influência que uma ou mais variáveis (explicativas) medidas em indivíduos ou objetos, têm sobre uma variável de interesse a que o nome de variável resposta. O modo como, em geral, o estatístico aborda tal problema é através do estudo de um modelo de regressão que relacione essa variável de interesse com as variáveis ditas explicativas.

O modelo linear normal, "criado" no início do século XIX, por Legendre e Gauss, dominou a modelagem estatística até meados do século XX, embora vários modelos não lineares ou não normais tenham sido desenvolvidos para fazer face a situações que não eram adequadamente explicadas pelo modelo linear normal. São exemplo disso, tal como referem McCullagh e Nelder (1989) e Lindsey (1997), o modelo complementar log-log para ensaios de diluição (Fisher, 1922), os modelos probit (Bliss, 1935) e logit (Berkson, 1944; Dyke and Patterson, 1952; Rasch, 1960) para proporções, os modelos log-lineares para dados de contagens (Birch, 1963), os modelos de regressão para análise de sobrevivência (Feigl e Zelen, 1965; Zippin e Armitage, 1966; Glasser, 1967).

Todos os modelos anteriormente descritos apresentam uma estrutura de regressão linear e têm em comum, o fator da variável resposta seguir uma distribuição dentro de uma família de distribuições com propriedades muito específicas: a família exponencial. Os Modelos Lineares Generalizados (MLG) introduzidos por Nelder e Wedderburn (1972) correspondem a uma síntese destes e de outros modelos, vindo assim unificar, tanto do ponto de vista teórico como conceitual, a teoria da modelagem estatística até então desenvolvida. Desde sua formulação os MLG

tem crescido em número de estudos teóricos e aplicações, sendo atualmente uma das principais áreas de estudo da estatística.

Neste trabalho, utilizou-se análises de dados binários, utilizando uma adaptação dos dados de Little (1978), distribuição de 1.607 mulheres casadas e aptas a terem filhos. Estas foram entrevistadas na pesquisa Fiji Fertilidade no ano de 1975. Em que a partir destes dados será utilizado a distribuição binomial sob a nomenclatura dos MLG e se fará a comparação entre as funções de ligação (probit e logit).



## 2 Revisão de Literatura

Nesta seção serão abordados os principais métodos estatísticos que servirão de base para o entendimento dos Modelos Lineares Generalizados via distribuição Binomial.

### 2.1 Família Exponencial

Segundo Cordeiro e Demétrio (2010) muitas das distribuições são membros da chamada classe ou família exponencial. Assim, por exemplo, as distribuições normal, binomial, binomial negativa, gama, poisson, normal inversa, etc. Esta classe de famílias de distribuições foi descoberta independentemente por Koopman, Pitman e Darrois através do estudo de propriedades de suficiência estatística. Posteriormente, muitos outros aspectos dessas famílias foram descobertos e tornaram-se importantes na teoria moderna da estatística a partir do trabalho pioneiro de Nelder e Wederburn (1972) que definiram os modelos lineares generalizados.

Os modelos lineares generalizados pressupõem que a variável resposta tenha uma distribuição pertencente a uma família particular, a família exponencial. Essa definição adequa-se a (Cox e Hinkley, 1974). Uma definição mais geral de família exponencial k-paramétrica e suas propriedades.

Diz-se que uma variável aleatória  $Y$  tem distribuição pertencente à família exponencial de dispersão (ou simplesmente família exponencial) se a sua função densidade de probabilidade (f.d.p.) ou função massa de probabilidade (f.m.p.) poder ser escrita na forma da família exponencial uniparamétrica:

$$f(y; \theta) = h(y) \exp[n(\theta)t(y) - b(\theta)] \quad (2.1)$$

com que as funções  $n(\theta)$ ,  $b(\theta)$ ,  $t(y)$  e  $h(y)$  assumem valores  $np$  em subconjunto dos reais. As funções  $n(\theta)$ ,  $b(\theta)$  e  $t(y)$  não são únicas. Por exemplo,  $n(\theta)$  pode ser multiplicada por uma constante  $k$  e  $t(y)$  pode ser dividido pela mesma constante.

Inúmeras distribuições podem ser representadas na forma (2.1), tais como: poisson, binomial, rayleigh, normal, gama e normal inversa (sendo as três últimas com a proposição de que pelo menos um dos parâmetros seja conhecido). O artigo de Cordeiro et al. (1995), apresentou 24 distribuições na forma (2.1). A base da família exponencial (2.1), isto é,  $\{y; f(y; \theta) > 0\}$ , não pode depender de  $\theta$ . Portanto, a distribuição uniforme em  $(0, \theta)$  não é um modelo da família exponencial. Pelo teorema da fatoração de Neyman-Fisher, a estatística  $t(y)$  é suficiente para  $\theta$ .

O elemento aleatório de um modelo linear generalizado é determinado a partir da família exponencial uniparamétrica na forma canônica  $f(y; \theta) = h(y)exp[\theta(y) - b(\theta)]$  sendo introduzido um parâmetro  $\phi > 0$  que é uma medida de dispersão da distribuição. Nelder e Wedderburn (1972) ao proporem esta teoria, conseguiram incorporar distribuições biparamétricas no componente aleatório do modelo, Isto é,

$$f(y; \theta, \phi) = exp\{\phi^{-1}[y\theta - b(\theta)] + c(y, \theta)\}, \quad (2.2)$$

em que  $b(\cdot)$  e  $c(\cdot)$  são funções conhecidas. Quando  $\phi$  é conhecido, a família de distribuições (2.2) será idêntica à família exponencial na forma canônica. O valor esperado e a variância de  $Y$ , com distribuição na família (2.2), é definido por

$$E(Y) = \mu = b'(\theta) \quad e \quad Var(Y) = \phi b''(\theta).$$

A partir da expressão da variância,  $\phi$  será um parâmetro de **dispersão** do modelo sendo seu inverso ( $\phi^{-1}$ ), uma medida de **precisão**. A função que melhor relaciona o parâmetro canônico  $\theta$  com a média  $\mu$  (inversa da função  $b'(\cdot)$ ) será denotado por  $\theta = q(\mu)$ . A função da média  $\mu$  na variância é representada por  $b''(\theta) = V(\mu)$ . Define-se  $V(\mu)$  como sendo **função de variância**. Pode-se observar que o parâmetro canônico é obtido de  $\theta = \int V^{-1}(\mu)d\mu$ , pois  $V(\mu) = \frac{d\mu}{d\theta}$ . A família

de distribuições (2.2) admite incorporar distribuições que apresentam assimetria de natureza discreta ou contínua e com suportes que são limitados a intervalos do conjunto dos reais.

Morris, 1982 demonstra que podem existir apenas seis distribuições na família (2.2) cuja função de variância é uma função, no máximo, quadrática da média. Essas distribuições são normal ( $V = 1$ ), gama ( $V = \mu^2$ ), binomial ( $V = \mu(1 - \mu)$ ), poisson ( $V = \mu$ ), binomial negativa ( $V = \mu + \frac{\mu^2}{k}$ ) e a sexta, chamada secante hiperbólica generalizada ( $V = 1 + \mu^2$ ).

A função geradora de momentos (f.g.m.) da família (2.2) é dada por

$$M(t; \theta, \phi) = E(e^{tY}) = \exp\{\phi^{-1}[(\phi t + \theta) - b(\theta)]\}. \quad (2.3)$$

Para este caso será feita apenas o caso de variáveis aleatórias contínuas. Para o caso discreto, basta substituir a integral pelo somatório. A função geradora de cumulantes (f.g.c.) correspondente é,

$$\varphi(t; \theta, \phi) = \log[M(t; \theta, \phi)] = \phi^{-1}[b(\phi t + \theta) - b(\theta)]. \quad (2.4)$$

Na f.g.c. tem um papel fundamental, mas que a f.g.m. na estatística, pois uma grande parte da teoria assintótica vai depender de suas propriedades. Derivando-se (2.4), sucessivamente, em relação a  $t$ , vem

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

em que  $b^{(r)}(\cdot)$  vai indicar a derivada de  $r$ -ésima ordem de  $b(\cdot)$  em relação a  $t$ . Para  $t = 0$ , obtendo-se o  $r$ -ésimo cumulante da família (2.2) como

$$K_r = \phi^{r-1} b^{(r)}(\theta). \quad (2.5)$$

Tendo sido enfatizado anteriormente, pode-se agora deduzir, a partir de (2.5), o valor esperado  $k_1 = \mu$  e a variância  $k_2$  da família (2.2) para  $r = 1$  e 2, respectivamente. Tem-se que  $k_1 = \mu = b'(\theta)$  e  $k_2 = \phi b''(\theta) = \phi \frac{d\mu}{d\theta}$ .

Na equação (2.4) pode-se notar que existe uma relação interessante de recorrência entre os cumulantes da família (2.2), isto por que,  $k_{r+1} = \phi \frac{dk_r}{d\theta}$  para  $r = 1, 2, \dots$ , fundamental para obtenção de propriedades assintóticas dos estimadores de máxima verossimilhança nos modelos lineares generalizados.

Podem-se, alternativamente, concluir essas expressões, simplesmente usando as propriedades da função escore. Ou seja  $l = l(\theta, \phi) = \log f(y; \theta, \phi)$  o logaritmo da função de máxima verossimilhança correspondente a uma única observação em (2.2). Tem-se

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

consequentemente,

$$E(\mathbf{U}) = \phi^{-1}[E(Y) - b'(\theta)] = 0 \text{ que sugere } E(Y) = b'(\theta)$$

logo

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

Portanto,

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

Conforme Bolfarine e Sandoval (2001); Casella e Berger (2011); Cordeiro e Demétrio (2010), uma estatística  $T = T(Y)$  é suficiente para  $\theta$  quando possuir toda informação acerca deste parâmetro e estiver contida na amostra  $Y$ . Se  $T$  for suficiente para  $\theta$ , então, a distribuição condicional de  $Y$  dada uma estatística  $T(Y)$  será independente de  $\theta$ , isto é,

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

Para o discernimento da fatoração tem-se uma forma conveniente de caracterizar uma estatística suficiente. Uma condição necessária e suficiente para que  $T$  seja suficiente para um parâmetro  $\theta$  é que a f.d.p.  $f_Y(y; \theta)$  possa ser decomposta como

$$f_Y(y; \theta) = h(y)g(t, y),$$

ainda que  $t = T(y)$  e  $h(y)$  não vão depender de  $\theta$ . Já para o caso discreto, a f.d.p. seria substituída pela função de probabilidade.

Seja  $Y_1, \dots, Y_n$  uma amostra aleatória (a.a.) de uma distribuição que pertence à família (2.2). A distribuição conjunta de  $Y_1, \dots, Y_n$  é definida

$$f(y; \theta, \phi) = \prod_{i=1}^n f(y_i; \theta, \phi) = \prod_{i=1}^n \exp\{\phi^{-1}[y_i\theta - b(\theta)] + c(y_i, \phi)\}$$

$$f(y; \theta, \phi) = \exp\{\phi^{-1}[\theta \sum_{i=1}^n y_i - nb(\theta)]\} \exp[\sum_{i=1}^n c(y_i, \phi)].$$

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

$$f(y; \theta, \phi) = g(t, \theta)h(h_1, \dots, h_n),$$

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

Sendo assim pode-se observar que se uma distribuição pertence à família exponencial, então, pode-se dizer que existe uma estatística suficiente. Na realidade, usando-se o Teorema de Lehmann-Scheffé (MENDENHALL et al., 1981) mostra-se que  $T = \sum_{i=1}^n Y_i$  vai ser uma estatística suficiente minimal.

## 2.2 Modelos Lineares Generalizados

Os Modelos Lineares Generalizados foram introduzidos no início dos anos 70, tendo um impacto muito grande no desenvolvimento da estatística aplicada. No

princípio o seu uso esteve confinado a um grupo restrito de investigadores, dada a falta de bibliografia acessível e a complexidade inicial do GLIM, primeiro software totalmente dirigido a aplicação da metodologia. Este avanço só foi possível a partir do momento em que o software existente passou a ser um pouco mais “amigável”. Atualmente, a maioria dos pacotes estatísticos de maior expansão já contem módulos adequados ao estudo destes modelos. Pode pois dizer-se que, neste momento, o conhecimento adequado da metodologia dos MLG é imprescindível para qualquer indivíduo que utilize métodos estatísticos.

Os modelos lineares generalizados (MLG) foram estabelecidos por John Nelder e Robert Wedberburn em 1972, estes autores mostram várias técnicas como uma forma de unificar vários modelos estatísticos, frequentemente estudadas separadamente, formulando-as, de uma maneira unificada, em uma classe de modelos de regressão. Tais modelos envolvem uma variável resposta univariada, variáveis explanatórias e uma amostra aleatória de  $n$  observações independentes do modelo. Possui uma distribuição pertencente à família de distribuições que engloba a normal, gama e normal inversa para todos os dados contínuos; binomial para proporções; poisson e binomial negativa para contagens;

A definição dos modelos lineares generalizados pode ser utilizada quando se tem uma única variável aleatória  $Y$  associada a um conjunto de variáveis explanatórias  $x_1, \dots, x_p$ . Para uma amostra de  $n$  observações  $(y_i, x_i)$  em que  $x_i = (x_{i1}, \dots, x_{ip})^T$  vai ser um vetor coluna de variáveis explicativas, o MLG envolve os três componentes: Componente aleatório, componente sistemático e função de ligação.

O componente aleatório é representado por um conjunto de variáveis aleatórias independentes  $Y_1, \dots, Y_n$  provenientes de uma mesma distribuição que faz parte da família de distribuições (2.2) com médias  $\mu_1, \dots, \mu_n$ , ou seja,

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

sendo  $\phi > 0$  um parâmetro de dispersão e o parâmetro  $\theta_i$  denominado parâmetro

canônico. Então, a f.d.p. de  $Y_i$  vai ser dada por

$$f(y_i; \theta_i, \phi) = \exp\{\phi^{-1}[y_i\theta_i - b(\theta_i)] + c(y_i, \phi)\} \quad (2.6)$$

sendo  $b(\cdot)$  e  $c(\cdot)$  funções conhecidas.

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

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

$$\theta_i = \int V_i^{-1} \partial \mu_i = q(\mu_i), \quad (2.7)$$

sendo  $q(\mu_i)$  uma função conhecida da média  $\mu_i$ . Dada uma relação funcional para a função de variância  $V(\mu)$ , o parâmetro canônico será obtido da equação (2.7) e a distribuição fica determinada na família exponencial (2.6). A importância da família (2.7) na teoria dos MLG's é que ela permite incorporar dados que exibem assimetria, dados de natureza discreta ou contínua e dados que são restritos a um intervalo do conjunto dos reais, como o intervalo  $(0, 1)$ .

No componente sistemático variáveis explicativas que entram na forma de uma soma linear de seus efeitos

$$\eta_i = \sum_{r=1}^p x_{ir} \beta_r = x_i^T \beta \quad \text{ou} \quad \eta_i = X \beta \quad (2.8)$$

sendo  $X = (x_1, \dots, x_n)^T$  a matriz do modelo,  $\beta = (\beta_1, \dots, \beta_p)^T$  o vetor de parâmetros e  $\eta = (\eta_1, \dots, \eta_n)^T$  o preditor linear. Se um parâmetro tem valor conhecido, o termo correspondente na estrutura linear é chamado *offset* que é um valor conhecido na regressão.

A função de ligação é uma função que relaciona o componente aleatório ao

componente sistemático, vincula a média ao preditor linear, isto é,

$$\eta_i = g(\mu_i), \quad (2.9)$$

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

Assim, vê-se que para a especificação do modelo, os parâmetros  $\theta_i$  da família de distribuições (2.6) não são de interesse direto (pois há um para cada observação) mas sim um conjunto menor de parâmetros  $\beta_1, \dots, \beta_p$  tais que uma combinação linear dos  $\beta'$  seja igual a alguma função do valor esperado de  $Y_i$ . Como o parâmetro natural  $\theta_i$  é uma função unívoca da média  $\mu_i$ , o mesmo pode ser expresso por uma função de ligação em termos desse parâmetro, isto é,  $\eta_i = g(q^{-1}(\theta_i))$  :

Nesses termos, MLG é definido por uma distribuição da família (2.6), uma estrutura linear (2.8) e uma função de ligação (2.9). Por exemplo, quando  $\theta = \mu$  e a função de ligação é linear, obtém-se o modelo clássico de regressão como um caso particular. Os modelos log-lineares são deduzidos supondo  $\theta = \log(\mu)$  com função de ligação logarítmica  $\text{Log}(\mu) = \eta$ .

As funções de ligação usuais são: potência  $\eta = \mu^\lambda$  em que  $\lambda$  é um número real, logística  $\eta = \log\left[\frac{\mu}{m-\mu}\right]$ , probito  $\eta = \phi^{-1}\left(\frac{\mu}{m}\right)$  sendo  $\phi(\cdot)$  a função de distribuição acumulada (f.d.a.) da distribuição normal padrão e a complemento  $\log - \log \eta = \log\left[-\log\left(1 - \frac{\mu}{m}\right)\right]$ , em que  $m$  é o número de ensaios independentes.

Aranda-Ordaz (1981) propôs uma família de funções de ligação para análise de dados na forma de proporções. Outra família importante de funções de ligação, principalmente para dados com média positiva, é a família potência (Cordeiro e Demétrio, 2010).

No modelo binomial a distribuição foi deduzida por James Bernoulli em (1713), embora tenha sido encontrada anteriormente em trabalhos de Pascal.

Suponha que  $Y = \eta\pi$  tenha distribuição binomial  $B(\eta; \pi)$ , com função de probabilidade, sendo que  $\pi$  representa a proporção de sucessos em  $m$  ensaios independentes com probabilidade de sucesso  $\pi$ .



O modelo binomial é usado, principalmente, no estudo de dados na forma de proporções, como nos casos da análise probito, logística (ou "logit") (ASHTON, 1972) e complemento log-log (Fisher, 1922), e na análise de dados binários, como na regressão logística linear (COX, 1970).

## 2.3 Estimação

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

As distribuições a serem assumidas para escala de tolerância e correspondentes funções de ligação devem ser capazes de transformar o intervalo  $(0, 1)$  em  $(-\infty, \infty)$ . Neste sentido, as distribuições logística, normal padrão e Gumbel (ou distribuição de valor extremo) para tolerância e suas correspondentes funções de ligação denominadas logito, probito e complemento log-log são apropriadas para o modelo binomial. A função de ligação complemento log-log ( $\log[-\log(1 - \mu)]$ ) pode ser mais indicada para valores de  $\mu$  próximos de zero.

Existe pouca diferença entre as distribuições normal e logística para tolerância ou escala contínua (CORDEIRO, 1986). Entretanto, a função de ligação logito é mais usada em dados binomiais pois refere-se a uma ligação canônica ou natural, tendo uma interpretação mais simples.

Considerando-se a função de ligação logito e a variável sobrevivência, em que os indivíduos vivos recebe o valor 1 e os mortos valor 0, tem-se:  $\mu_i = P(Y_i = 1) =$  probabilidade de um indivíduo sobreviver;  $g(\mu_i) = \eta_i = \log\left[\frac{\mu_i}{1-\mu_i}\right]$ ;  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} =$

modelo linear imposto ao logito.

A variância residual, ou seja, a variância das observações dado  $\eta_i$  (ou  $\mu_i$ ), advém da amostragem binomial, ou seja:

$$\text{Var}(Y_i|\mu_i) = \mu_i(1 - \mu_i).$$

Se um MLG for definido pelas expressões (2.6), (2.8) e (2.9) e supondo-se que os dados  $\mathbf{y} = (y_1, \dots, y_n)^T$ : O logaritmo da função de máxima verossimilhança como função apenas de  $\boldsymbol{\beta}$  (considerando-se o parâmetro de dispersão  $\phi$  conhecido) dado o vetor  $\mathbf{y}$  é definido por  $l(\boldsymbol{\beta}) = l(\boldsymbol{\beta}, \mathbf{y})$  e usando-se a expressão (2.6) tem-se

$$l(\boldsymbol{\beta}) = \phi^{-1} \sum_{i=1}^n [y_i \theta_i - b(\beta_i)] + \sum_{i=1}^n c(y_i, \phi), \quad (2.10)$$

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

Deve-se, primeiramente, estimar o vetor de parâmetros  $\boldsymbol{\beta}$  para depois calcular as estimativas do vetor das médias  $\boldsymbol{\mu}$  e do vetor dos parâmetros  $\boldsymbol{\theta}$  pelas relações funcionais  $\mu_i = g^{-1}(x_i^T \boldsymbol{\beta})$  e  $\theta_i = q(\mu_i)$ :

Um caso importante dos MLG surge quando o vetor de parâmetros canônicos  $\boldsymbol{\theta}$  da família (2.6) e o vetor de preditores lineares  $\boldsymbol{\eta}$  em (2.8) são iguais, conduzindo-se às funções de ligação canônicas. Tem-se,  $\theta_i = \eta_i = \sum_{r=1}^p x_{ir} \beta_r$  para  $i = 1, \dots, n$ : As estatísticas  $S_r = \sum_{i=1}^n x_{ir} Y_i$  para  $r = 1, \dots, p$  são suficientes para os parâmetros  $\beta_1, \dots, \beta_p$  e têm dimensão máxima  $p$ . Sejam  $s_r = \sum_{i=1}^n x_{ir} y_i$  as realizações de  $S_r$ ,  $r = 1, \dots, p$ . Então, (2.10) pode ser escrita na forma:

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

e, portanto,  $l(\boldsymbol{\beta})$  tem a seguinte decomposição  $l(\boldsymbol{\beta}) = l_1(\mathbf{s}, \boldsymbol{\beta}) + l_2(\mathbf{y})$ , em que  $l_1(\mathbf{s}, \boldsymbol{\beta}) = \phi^{-1} \sum_{r=1}^p s_r \beta_r - \phi^{-1} \sum_{i=1}^n b(\sum_{r=1}^p x_{ir} \beta_r)$  e  $l_2(\mathbf{y}) = \sum_{i=1}^n c(y_i, \phi)$ .

As funções de ligação canônicas produzem propriedades estatísticas de interesse para o modelo, tais como, suficiência, facilidade de cálculo, unicidade das estimativas de máxima verossimilhança e, em alguns casos, interpretação simples e, a princípio, pode-se trabalhar com ela quando não existirem indicativos de outra preferível. Entretanto, não existe razão para se considerarem sempre os efeitos sistemáticos como aditivos na escala dada pela função de ligação canônica.

Inicialmente, considera-se esse trinômio fixo para se obter uma descrição adequada dos dados através das estimativas dos parâmetros do modelo. Muitos métodos podem ser usados para estimar os parâmetros  $\beta' s$ ; inclusive o qui-quadrado máximo, o Bayesiano e a estimação-M. O último inclui o método de máxima verossimilhança (MV) que tem muitas propriedades ótimas, tais como, consistência e eficiência assintótica.

Considerado-se apenas o método de (MV) para estimar os parâmetros lineares  $\beta_1, \dots, \beta_p$  do modelo. O vetor escore é formado pelas derivadas parciais de primeira ordem do logaritmo da função de verossimilhança. Da expressão (2.10) pode-se calcular, pela regra da cadeia, o vetor escore  $U(\beta) = \frac{\partial l(\beta)}{\partial \beta}$  de dimensão  $p$ , com elemento típico  $U_r = \frac{\partial l(\beta)}{\partial \beta_r} = \sum_{i=1}^n \frac{d l_i}{d \theta_i} \frac{d \theta_i}{d \mu_i} \frac{d \mu_i}{d \eta_i} \frac{\partial \eta_i}{\partial \beta_r}$ , pois

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

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

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

$$n_i = \sum_{r=1}^p x_{ri} \beta_r$$

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

$$U_r = \phi^{-1} \sum_{i=1}^n (y_i - \mu_i) \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} x_{ir} \quad (2.11)$$

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

O método iterativo de Newton-Raphson para a solução de uma equação  $f(y) = 0$  é baseado na aproximação de Taylor para a função  $f(y)$  na vizinhança do ponto  $y_0$ , ou seja,

$$f(y) = f(y_0) + (y - y_0)f'(y_0) = 0,$$

obtendo-se

$$y = y_0 - \frac{f(y_0)}{f'(y_0)},$$

ou, de uma forma mais geral,

$$y^{(v-1)} = y^{(v)} - \frac{f(y^{(v)})}{f'(y^{(v)})},$$

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

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

$$\beta^{(m+1)} = \beta^{(m)} + (\mathbf{K}^{(m)})^{-1} \mathbf{U}^{(m)} \quad (2.12)$$

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

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

que é a matriz de covariâncias dos  $\mathbf{U}'_r\mathbf{S}$ :

Multiplicando-se ambos os membros de (2.12) por  $\mathbf{K}^{(m)}$ , tem-se

$$\mathbf{K}^{(m)}\beta^{(m+1)} = \mathbf{K}^{(m)}\beta^{(m)} + \mathbf{U}^{(m)}. \quad (2.13)$$

Substituindo  $\mathbf{K}$  e  $\mathbf{U}$  em (2.13) e eliminando  $\phi$ , tem-se

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

ou, ainda,

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

Define-se a variável dependente ajustada  $z = \eta + G(y, \mu)$ : Logo,

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

ou

$$\beta^{(m+1)} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} z^{(m)}. \quad (2.14)$$

A equação matricial (2.14) é válida para qualquer MLG e mostra que a solução das equações de MV equivale a calcular repetidamente uma regressão linear ponderada de uma variável dependente ajustada  $z$  sobre a matriz  $X$  usando uma matriz de pesos  $W$  que se modifica no processo iterativo.

A demonstração de (2.14), em generalidade, foi dada por Nelder e Wedder-Burn (1972). Eles generalizaram procedimentos iterativos obtidos para casos especiais dos MLG: probit (FISHER, 1935), log-lineares (HABERMAN, 1970) e log-lineares (COX, 1972).

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

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

Em pequenas amostras, a Equação (2.14) pode divergir. O número de iterações até convergência depende inteiramente do valor inicial arbitrado para  $\hat{\beta}$ , embora, geralmente, o algoritmo convirja rapidamente. A desvantagem do método tradicional de Newton-Raphson com o uso da matriz observada de derivadas de segunda ordem é que, normalmente, não converge para determinados valores iniciais.

Nas funções de ligação canônicas  $\mathbf{w} = \mathbf{V} = \frac{d\mu}{d\eta}$  que produzem os modelos denominados canônicos, as equações de MV têm a seguinte forma, facilmente deduzidas de (2.11),

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

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

$$\mathbf{X}^T \mathbf{y} = \mathbf{X}^T \hat{\boldsymbol{\mu}}. \quad (2.15)$$

Sendo assim, as estimativas de MV dos  $\beta'$ s são únicas. Sendo  $\mathbf{S} = (S_1, \dots, S_p)^T$  o vetor de estatísticas suficientes definidas por  $S_r = \sum_{i=1}^n x_{ir} Y_i$ , e  $\mathbf{s} = (s_1, \dots, s_p)^T$  os seus valores amostrais, a Equação (2.15) podem ser expressos por

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

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

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

O algoritmo (2.14) pode ser usado para ajustar inúmeros outros modelos, como aqueles baseados na família exponencial (2.1) que estão descritos por Cordeiro et al. (1995), bastando identificar as funções de variância e de ligação.

Com a obtenção da EMV  $\widehat{\beta}$  em (2.15), podem-se calcular as EMV dos preditores lineares  $\widehat{\eta} = \mathbf{X}\widehat{\beta}$  e das médias  $\widehat{\mu} = g^{-1}(\widehat{\eta})$ . A EMV do vetor  $\theta$  de parâmetros canônicos é, simplesmente, igual a  $\widehat{\theta} = q(\widehat{\mu})$ :

A inversa da matriz de informação estimada em  $\widehat{\beta}$  representa a estrutura de covariância assintótica de  $\beta$ , isto é, a matriz de covariância de  $\beta$  quando  $n \rightarrow 1$ : Logo, a matriz de covariância de  $\beta$  é estimada por

$$\widehat{Cov}(\beta) = \phi(\mathbf{X}^T \widehat{\mathbf{W}}^T \mathbf{X})^{-1} \quad (2.16)$$

em que  $\widehat{\mathbf{W}}$  é a matriz de pesos  $\mathbf{W}$  avaliada em  $\widehat{\beta}$ .

Intervalos de confiança assintóticos para os parâmetros  $\beta's$  podem ser deduzidos da aproximação (2.16). Observa-se que o parâmetro de dispersão  $\phi$  é um fator multiplicativo na matriz de covariância assintótica de  $\widehat{\beta}$ .

A estrutura da covariância assintótica das EMV dos preditores lineares em  $\widehat{\eta}$  é obtida diretamente de  $Cov(\widehat{\eta}) = \mathbf{X}Cov(\widehat{\beta})\mathbf{X}^T$ : Logo,

$$\widehat{Cov}(\widehat{\eta}) = \phi \mathbf{X}(\mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T. \quad (2.17)$$

A matriz  $\mathbf{Z} = \{z_{ij}\} = \mathbf{X}(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T$  que aparece em (2.17) tem um papel importante na teoria assintótica dos MLG (CORDEIRO, 1983; CORDEIRO e MCCULLAGH, 1991). Essa matriz surge no valor esperado da função desvio até termos de ordem  $O(n^{-1})$  e no valor esperado da estimativa  $\widehat{\eta}$  até essa ordem.

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

$$Cov(\widehat{\eta}) = \mathbf{G}^{-1} cov(\widehat{\eta}) \mathbf{G}^{-1} \quad (2.18)$$

enfatizando que a matriz diagonal  $\mathbf{G} = diag \frac{d\eta_i}{d\mu_i}$ .

Expandindo a função suporte  $l = l(\beta)$ , em série multivariada de Taylor ao

redor de  $\widehat{\beta}$  e notando que  $U(\widehat{\beta}) = 0$ , obtém-se, aproximadamente,

$$\widehat{\mathbf{1}} - \mathbf{1} \doteq \frac{1}{2}(\beta - \widehat{\beta})^T \widehat{\mathbf{J}}(\beta - \widehat{\beta}) \quad (2.19)$$

A forma quadrática do lado direito de (2.19) aproxima a superfície suporte por um parabolóide, passando pelo seu ponto de máximo, com a mesma curvatura esférica da superfície nesse ponto.

A Equação (2.19) implica que a função de verossimilhança  $L = L(\beta)$  num ponto qualquer  $\beta$  segue, aproximadamente, a expressão

$$L \doteq \exp\left[-\frac{1}{2}(\beta - \widehat{\beta})^T \widehat{\mathbf{J}}(\beta - \widehat{\beta})\right] \quad (2.20)$$

em que  $\widehat{L}$  é a função de verossimilhança avaliada em  $\widehat{\beta}$ , que representa a forma da curva normal multivariada com média  $\widehat{\beta}$  e estrutura de covariância igual a  $\widehat{\mathbf{J}}^{-1}$ .

A fórmula (2.20) apresenta decomposição da função de verossimilhança, pelo menos para  $n$  grande, estabelecendo, pelo teorema da fatoração, a suficiência assintótica da EMV. Conclui-se que, embora as EMV não sejam necessariamente suficientes para os parâmetros do modelo, essa suficiência será alcançada quando a dimensão do vetor de observações tender para infinito.

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

$$\widehat{\beta} - \beta \doteq \mathbf{J}^{-1}\mathbf{U} \quad (2.21)$$

O método de Newton-Raphson, de cálculo da EMV consiste em usar a Equação (2.21) iterativamente. Obtém-se uma nova estimativa  $\widehat{\beta}^{(m+1)}$  a partir de uma



anterior  $\beta^{(m)}$  por meio de

$$\beta^{(m+1)} = \beta^{(m)} + \mathbf{J}^{(m)-1} \mathbf{U}^{(m)} \quad (2.22)$$

em que quantidades avaliadas na  $m$ -ésima iteração do procedimento iterativo são indicadas com o superescrito  $(m)$ . O processo é, então, repetido a partir de  $\beta^{(m+1)}$  até a distância entre  $\beta^{(m+1)}$  e  $\beta^{(m)}$  se tornar desprezível ou menor do que uma quantidade pequena especificada.

A expressão (2.21) tem uma forma alternativa equivalente, assintoticamente, pois pela lei dos grandes números  $\mathbf{J}$  deve convergir para  $\mathbf{K}$  quando  $n \rightarrow \infty$ . Assim, substituindo a informação observada em (2.21) pela esperada, obtém-se a aproximação

$$\hat{\beta} - \beta \doteq \mathbf{K}^{-1} \mathbf{U}. \quad (2.23)$$

O procedimento iterativo baseado em (2.23) é denominado método escore de Fisher para parâmetros, isto é,  $\beta^{(m+1)} = \beta^{(m)} + \mathbf{K}^{(m)-1} \mathbf{U}^{(m)}$ , como foi explicitado na equação (2.12). O aspecto mais trabalhoso dos dois esquemas iterativos é a inversão das matrizes  $\mathbf{J}$  e  $\mathbf{K}$ . Ambos os procedimentos são muito sensíveis em relação à estimativa inicial  $\beta^{(1)}$ . Se o vetor  $\beta^{(1)}$  for uma estimativa consistente, ambos os métodos convergirão em apenas um passo para uma estimativa eficiente, assintoticamente.

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

$$\beta^{(m-1)} = \beta^{(m)} - \mathbf{s}^{(m)} \mathbf{Q}^{(m)} \mathbf{U}^{(m)}, \quad (2.24)$$

em que  $\mathbf{s}^{(m)}$  é um escalar,  $\mathbf{Q}^{(m)}$  é uma matriz quadrada que determina a direção

da mudança de  $\beta^{(m)}$  para  $\beta^{(m+1)}$  e  $\mathbf{U}^{(m)}$  é o vetor gradiente do logaritmo da função de verossimilhança  $l(\beta)$ , com todas essas quantidades variando no processo iterativo. Os algoritmos iniciam num ponto  $\beta^{(1)}$  e procedem, por meio da expressão  $\frac{\partial l(\theta; y)}{\partial \theta} = 0$ , para calcular aproximações sucessivas para a EMV  $\hat{\beta}$ . Nos procedimentos iterativos de Newton-Raphson e escore de Fisher,  $\mathbf{s}^{(m)}$  é igual a um, e a matriz de direção  $\mathbf{Q}^{(m)}$  é igual à inversa da matriz Hessiana e à inversa do valor esperado dessa matriz, respectivamente. Esses dois procedimentos devem ser iniciados a partir de uma estimativa consistente a fim de se garantir convergência para  $\hat{\beta}$ . A escolha do melhor algoritmo em (2.24) é função da geometria do modelo em consideração e, em geral, não existe um algoritmo superior aos demais em qualquer espectro amplo de problemas de estimação.

## 2.4 Distribuição Binomial

Considera-se um caso em que a resposta é  $y_i$  em binário, assumindo-se apenas dois valores que, por conveniência, é um código de um (1) ou zero (0). Por exemplo, poderíamos definir desta forma:

$$y_i = \begin{cases} 1, & \text{se a mulher usa anticoncepcionais,} \\ 0, & \text{caso contrário.} \end{cases}$$

Tem-se assim  $y_i$  como uma realização de um  $Y_i$  de variável aleatória que pode assumir os valores um e zero com probabilidades  $\pi_i$  e  $1 - \pi_i$ , respectivamente. A distribuição  $Y_i$  é chamada uma distribuição de Bernoulli com parâmetro  $\pi_i$ , e pode ser escrita de forma compacta como

$$P\{Y_i = y_i\} = \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

para  $y_i = 0, 1$ . Observe que, se  $y_i = 1$  obtem-se  $\pi_i$ , e se  $y_i = 0$  obtemos  $1 - \pi_i$ .

Verifica-se por cálculo direto que o valor esperado e variância de  $Y_i$  são

$$E(Y_i) = \mu_i = \pi_i, \text{ e}$$

$$VAR(Y_i) = \sigma_i^2 = \pi_i(1 - \pi_i).$$

Note-se que a média e variância dependem da probabilidade subjacente  $\pi_i$ . Qualquer fator que tenha um a probabilidade de ocorre não irá alterar apenas a média, mas também a variância das observações. Isto sugere que o modelo linear vai permiti os preditores para a média, e assumem a variância constante que não será adequada para a análise de dados binários.

Suponha agora que as unidades em estudo possam ser classificadas de acordo com os fatores de interesse em grupos de  $k$ , de tal maneira que todos os indivíduos de um grupo tenha valores idênticos de todas as covariáveis. Em nossa aplicação, as mulheres podem ser classificadas em 16 diferentes grupos em termos de idade, escolaridade e desejo para mais crianças. Deixe  $\eta_i$  denotar o número de observações no grupo  $i$ , e deixa  $y_i$  que denota o número de unidades que possuem o atributo de interesse no grupo  $i$ . Por exemplo,

$y_i$  = número de mulheres que usam contracepção no grupo  $i$ .

Define-se  $y_i$  como uma realização de uma  $(Y_i)$  variável aleatória que assume os valores  $0, 1, \dots, \eta_i$ . Se as observações  $\eta_i$  em cada grupo são independentes, e eles têm a mesma probabilidade  $\pi_i$  de ter o atributo de interesse, em seguida, a distribuição de  $Y_i$  é binomial com parâmetros  $\pi_i$  e  $\eta_i$ , pode ser escrita comoque escrever

$$Y_i \sim B(\eta_i, \pi_i)$$

A função de distribuição de probabilidade de  $Y_i$  é dada por

$$P\{Y_i = y_i\} \binom{\eta_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{\eta_i - y_i}$$

para  $y_i = 0, 1, \dots, n_i$ . Aqui  $\pi_i^{Y_i}(1 - \pi_i)^{\eta_i - Y_i}$  é a probabilidade de se obter  $y_i$  sucessos e fracassos  $\eta_i - Y_i$  em alguma ordem tal especifica, e a combinação é o número de maneiras de obter sucessos de  $y_i$  em ensaios  $\eta_i$ .

A média e a variância de  $Y_i$  pode ser definido como

$$E(Y_i) = \mu_i = \eta_i \pi_i, \text{ e}$$

$$VAR(Y_i) = \sigma_i^2 = \eta_i \pi_i (1 - \pi_i).$$

A maneira mais fácil de obter este resultado é o seguinte. Seja  $Y_{ij}$  um indicador de variável que assume os valores um ou zero, se a unidade  $j$  no grupo  $i$  é um sucesso ou uma falha, respectivamente. Observa-se que  $Y_{ij}$  é uma variável aleatória com distribuição de Bernoulli. Pode-se escrever o número sucessos do grupo  $i$  como uma soma das variáveis indicadoras individuais, para  $Y_i = \sum_j^n Y_{ij}$ . A média de  $Y_i$  é, em seguida, a soma dos valores médios individuais, e pela independência, a sua variação é a soma dos desvios individuais. Observe mais uma vez que a média e a variância dependem da probabilidade de base  $\pi_i$ . Qualquer fator desta probabilidade terá uma média como o desvio das observações.

## 2.5 Transformação Logit

A regressão logística analisa dados distribuídos binomialmente da forma,

$$Y_i \sim B(p_i, n_i), \text{ for } i = 1, \dots, m,$$

onde os números de ensaios de Bernoulli  $n_i$  são conhecidos e as probabilidades de êxito  $p_i$  são desconhecidas. Um exemplo desta distribuição é a percentagem de sementes ( $p_i$ ) que germinam depois de  $n_i$  serem plantadas.

O modelo é então obtido na base de que cada ensaio (valor de  $i$ ) e o conjunto de variáveis explicativas e independentes que possa informar acerca da probabilidade

final. Estas variáveis explicativas podem-se ver como um vector  $X_i$   $k$ -dimensional e o modelo toma então a forma

$$p_i = E\left(\frac{Y_i}{n_i} \mid X_i\right).$$

Os logits com probabilidades binomiais desconhecidas (*i.e.*, os logaritmos dos odds) são modelados como uma função linear dos  $X_i$ .

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i}.$$

Note-se que um elemento particular de  $X_i$  pode ser ajustado a 1 para todo o  $i$  obtendo-se um intercepto no modelo. Os parâmetros desconhecidos  $\beta_j$  são habitualmente estimados através de máxima verossimilhança.

A interpretação dos valores estimados do parâmetro  $\beta_j$  é similar aos efeitos aditivos em log odds ratio para uma unidade de mudança na  $j$ -ésima variável explicativa. No caso de uma variável explicativa dicotómica, por exemplo o género,  $e^\beta$  é o estimador de odds ratio de ter o resultado para, por exemplo, homens comparados com mulheres.

O modelo tem uma formulação equivalente dada por

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i})}}.$$

Esta forma funcional é habitualmente identificada como um perceptron de uma camada simples ou rede neuronal artificial de uma só camada. Uma rede neuronal de uma só camada calcula uma saída contínua em vez de uma função por troços. A derivada de  $p_i$  em relação a  $X = x_1 \dots x_k$  é calculada na forma geral:

$$y = \frac{1}{1 + e^{-f(X)}}$$

onde  $f(X)$  é uma função analítica em  $X$ . Com esta escolha, a rede de camada simples é idêntica ao modelo de regressão logística. Esta função tem uma derivada contínua, a qual permite ser usada na propagação para trás. Esta função também é preferida pois a sua derivada é facilmente calculável:

$$y' = y(1 - y) \frac{df}{dX},$$

## 2.6 Transformação Probit

O probit é um método para estimar dose críticas em ensaios de dose-resposta, discutido em detalhes em Bliss (1935). Um ensaio do tipo dose-resposta é aquele onde uma determinada droga é administrada em  $k$  diferentes doses (níveis),  $d_1, d_2, \dots, d_k$  em, respectivamente,  $m_1, m_2, \dots, m_k$  indivíduos, obtendo-se como resposta, após um período especificado,  $y_1, y_2, \dots, y_k$  indivíduos que mudam de estado (ocorrência de um sucesso, por exemplo, morte). Neste tipo de ensaio uma amostra com indivíduos de uma mesma espécie é selecionada. Tal amostra é dividida em  $k$  grupos, cada um com  $m_i$  indivíduos,  $i = 1, 2, \dots, k$ . Cada dose é aplicada a cada grupo e o número de sucessos por grupo (morte do indivíduo, por exemplo) é contado. Dessa maneira, a resposta para cada grupo é uma variável aleatória com distribuição binomial com parâmetros  $m_i$  e  $\pi_i$ , ou seja,  $Y_i \sim \text{Bin}(m_i, \pi_i)$ .

No método probit, tal variável será considerada no processo de modelagem como uma variável aleatória com distribuição normal com média  $\mu$  e variância  $\sigma^2$ , ou seja, se  $U$  é a variável aleatória que representa a Tolerância de cada indivíduo da população, tem-se que  $U \sim N(\mu, \sigma^2)$ . Se a dose  $d_i$  é aplicada à população toda e  $f_U(u)$  é a função de densidade de probabilidade para a distribuição de Tolerâncias, todo indivíduo cuja Tolerância é menor do que  $d_i$  responderá à droga (tratamento), e a probabilidade de que um indivíduo escolhido ao acaso responda à dose é dada por:

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

Assim,

$$\pi_i = P[Z \leq -\frac{\mu}{\sigma} + \frac{1}{\sigma}d_i] = P[Z \leq \beta_1 + \beta_2d_i],$$

com  $\beta_1 = -\frac{\mu}{\sigma}$  e  $\beta_2 = \frac{1}{\sigma}$ . Considerando  $\Phi(x)$  a distribuição acumulada da normal padrão no ponto  $x$ , tem-se:

$$\pi_i = \Phi(\beta_1 + \beta_2d_i) \Rightarrow \Phi^{-1}(\pi_i) = \beta_1 + \beta_2d_i.$$

O probit de  $\pi_i$  é definido como sendo:  $\text{probit}(\pi_i) = \Phi^{-1}(\pi_i) = \beta_1 + \beta_2d_i$ .

Encontrando-se estimativas para  $\beta_1$  e  $\beta_2$ , pode-se estimar os valores de  $\mu$  e  $\sigma$ . Para a estimação de dose letais elevadas ou baixas o método sofre de um problema de estabilidade em razão da calda da distribuição. Fisher (1935), sugere um método para contornar este problema: Considerando  $p_i$  a proporção de sucessos para a dose  $i$ , toma-se todos os pares  $(d_i, p_i)$  onde  $p_i \neq 0$  e  $p_i \neq 1$  e faz-se uma primeira estimativa de  $\beta_1$  e  $\beta_2$  por Mínimos Quadrados Ponderados com pesos  $w_i = \frac{z_i^2 m_i}{p_i q_i}$ . Encontradas estas estimativas, considera-se o probit para a dose  $d_j$  em que ocorreu  $p_j = 0$  por:

$$\text{probit}^*(p_j) = (\widehat{\beta}_1 + \widehat{\beta}_2 d_j) - \frac{\Phi(d_j)}{\phi(d_j)},$$

e, para o caso onde ocorreu  $p_j = 1$ :

$$\text{probit}^*(p_j) = (\widehat{\beta}_1 + \widehat{\beta}_2 d_j) - \frac{1 - \Phi(d_j)}{\phi(d_j)}$$

De posse destes valores, considera-se todos os valores iniciais juntamente com os anteriores e a regressão linear é refeita, calculando-se novas estimativas para  $\beta_1$  e  $\beta_2$ , que serão as estimativas finais do processo.

## 2.7 Teste de Hipóteses

Segundo os métodos de inferência nos modelos lineares generalizados baseiam-se, fundamentalmente, na teoria da máxima verossimilhança. De acordo com esta teoria, existem três estatísticas para testar hipóteses relativas aos parâmetros  $\beta'$ s, que são deduzidas de distribuições assintóticas de funções adequadas das estimativas dos  $\beta'$ s. E estas estatísticas são: Razão de Verossimilhança, Wald e Escore. Assintoticamente equivalentes e, sob  $H_0$  e para  $\phi$  conhecido, convergem para uma variável com distribuição  $\chi_p^2$ , sendo, porém, a razão de verossimilhanças, o teste uniformemente mais poderoso.

A razão de verossimilhanças para testar componentes do vetor pode ser obtida como uma diferença de desvios entre modelos encaixados. A estatística de Wald é baseada na distribuição normal assintótica de  $\hat{\beta}$ . A estatística escore é obtida da função escore.

Segundo Cordeiro (2006), dependendo da hipótese a ser testada, em particular, qualquer uma dessas três estatísticas pode ser a mais apropriada. Para hipóteses relativas a um único coeficiente  $\beta_j$ , a estatística de Wald é a mais utilizada. Para hipóteses relativas a vários coeficientes, a razão de verossimilhanças é, geralmente, preferida. Dentre outras, referências importantes são Silvey (1975, pág.108-122), Cordeiro (1986), Dobson (1990), Paula (2000) e McCulloch, Searle (2001).

### 2.7.1 Teste de Wald

Conforme o teste de Wald é uma situação de hipótese dada anteriormente, a estatística de Wald é definido por

$$\xi_w = [\hat{\beta} - \beta^0] \phi(\mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X})^T [\hat{\beta} - \beta^0],$$

onde  $\phi(\mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X})$  e a  $\widehat{VAR}(\hat{\beta})$ . Podemos observar que se o número de parâmetros do vetor  $\beta$  for igual a 1, então pode-se escrever o teste de Wald como sendo equivalente



ao teste  $t^2$  que é utilizado normalmente no modelo clássico de regressão

$$\xi_w = \frac{(\widehat{\beta} - \beta^0)(\widehat{\beta} - \beta^0)}{\widehat{VAR}(\widehat{\beta})}.$$

Segundo (PAULA, 2004), um problema com a estatística de Wald é a dependência de  $\xi_w$  com a parametrização usada, quando  $\eta(\beta)$  é não-linear em  $\beta$ , ou seja, duas formas diferentes e equivalentes para  $\eta(\beta)$ , podem levar a diferentes valores de  $\xi_w$ .

## 2.8 Validação

### 2.8.1 Modelo Completo ou Saturado

Consideremos o modelo linear generalizado sem a estrutura linear  $\eta = Z\beta$ , isto é com  $n$  parâmetros  $\mu_1, \dots, \mu_n$ , linearmente independentes, sendo a matriz do modelo a matriz identidade  $n * n$ . Este modelo atribui toda a variação dos dados à componente sistemática. Como as estimativas de máxima verosimilhança dos  $\mu_i$  são as próprias observações, isto é,  $\widehat{\mu}_i = y_i$ , o modelo ajusta-se exatamente, reproduzindo os próprios dados. Não oferece qualquer simplificação e, como tal, não tem interesse na interpretação do problema, já que não tem características importantes transmitidas pelos dados. Além disso tem pouca hipótese de ser um modelo adequado em réplicas do estudo. A sua consideração é contudo necessária na formulação da teoria da seleção de modelos.

## 2.9 Qualidade do Ajustamento

Podemos avaliar a qualidade do ajuste do modelo fazendo um gráfico do modelo ajustado, mas existem medidas, como a deviance, a estatística qui-quadrada de Pearson e o teste de Hosmer e Lemeshow que ajudam a verificar a qualidade do modelo.

Para os testes de Deviance e qui-quadrado de Pearson, precisamos identificar, dentre as observações, aquelas com os mesmos valores entre si para todos as covariáveis. Assim, seja  $\mathbf{X} = (X_1, X_2, \dots, X_p)$  o conjunto de variáveis explicativas do modelo. Seja também  $J$  o número distinto de valores que  $x$  assume na amostra (combinações de níveis diferentes das covariáveis, chamado de Covariate Patterns). O número de indivíduos na amostra com valores iguais  $\mathbf{x} = \mathbf{x}_j$  é denotado por:  $m_j$ . Já  $y_j$  é o número de indivíduos positivos, isto é,  $y = 1$ , em  $m_j$ . Conseqüentemente,  $\sum y_j = n_1$  que representa o total de indivíduos com  $y = 1$ .

Já para o teste de Hosmer e Lemeshow não é necessário a existência de combinações de valores das covariáveis.

A soma dos quadrados dos resíduos (SQE) da regressão linear, no modelo logístico é denominada deviance (D), definida como:

$$d(y_j, \hat{\pi}_j) = \pm \left[ 2 \left[ y_j \ln \left( \frac{y_j}{m_j \hat{\pi}_j} \right) + (m_j - y_j) \ln \left( \frac{(m_j - y_j)}{m_j \hat{\pi}_j (1 - \hat{\pi}_j)} \right) \right] \right]^{1/2}$$

Em que o sinal + ou - é o mesmo sinal que  $(y_j - m_j \hat{\pi}_j)$ . Além disso,  $m_j$  é o número de observações na Covariate Pattern  $j$ ,  $\hat{\pi}_j$  é a probabilidade ajustada dos indivíduos no grupo  $j$  e  $y_j$  o número de indivíduos em  $j$  com  $y = 1$ .

A estatística teste é:

$$D = \sum_{j=1}^J d(y_j, \hat{\pi}_j)^2.$$

A estatística D, sob a suposição que o modelo ajustado é correto, tem distribuição assintótica  $\chi^2$  com  $J - (p + 1)$  graus de liberdade.

Restrições do teste Qui-Quadrado de Pearson e Deviance quando os resultados assintóticos são baseados no fato em que  $n$  cresce mantendo os  $m_j$  pequenos (ou tornando os  $m_j$  pequenos), chamamos de n-assintótico.

Quando  $J < n$  é fixado e  $n$  cresce fazendo com que os  $m_j$  também cresçam, os resultados assintóticos baseados no  $m_j$  são chamados de m-assintótico.

Assim, quando  $J \approx n$ , sob n-assintótico, os p-valores calculados dos testes Deviance e qui-quadrado de Pearson, usando a distribuição  $\chi^2$  com  $(J - p - 1)$  graus de liberdade estão incorretos. Uma saída é criar agrupamentos que permite o uso de resultados m-assintóticos. Vale ressaltar que para a realização do teste de qualidade de ajuste de Hosmer e Lemeshow não é necessário existir Covariate Pattern.

### 3 Aplicação

Os dados utilizados neste trabalho são de uma adaptação de Little (1978), que mostra a distribuição de 1.607 mulheres atualmente casadas e em idade fértil, entrevistadas na pesquisa Fiji Fertilidade de 1975. A classificação por idade atual, nível de educação, o desejo de ter mais filhos, além do uso de anticoncepcionais.

Para a análise destes dados a variável resposta será o uso atual de contraceptivos. E as variáveis idade, classificada por;  $A = [< 25]$ ,  $B = [25 - 29]$ ,  $C = [30 - 39]$  e  $D = [40 - 49]$ , escolaridade,  $B = (\text{Baixo})$  e  $A = (\text{Alto})$  e o desejo de ter mais filhos - (DMF), denotado por  $S = (\text{Sim})$  e  $N = (\text{Não})$ . Na Tabela 1 tem-se as frequências observadas para cada uma das variáveis. Nesta análise observa-se o uso atual de contracepção como a resposta ou variável dependente de interesse. As variáveis idade, escolaridade e desejo de ter mais filhos como preditores. Pode-se observar que a resposta tem duas categorias: uso e não uso. Neste exemplo, todos os indicadores são tratados como categóricas, mas as técnicas a serem estudadas podem ser aplicadas de forma mais geral para ambos os fatores discretos e contínuos.

O conjunto de dados original inclui a data de nascimento do entrevistado e a data da entrevista em forma de mês/ano, por isso, é possível calcular a idade em anos solteiras, mas vamos usar as faixas etárias de dez anos para a conveniência. Da mesma forma, o estudo inclui informações sobre o mais alto nível de educação atingido e o número de anos completos a esse nível, então pode-se calcular concluído anos de escolaridade, mas será trabalhado com uma simples distinção entre primário ou menos e fundamental ou mais.

Por fim, desejo para mais crianças é medida como uma simples dicotomia codificado em (sim) ou (não), e portanto, é, naturalmente, uma variável categórica. O fato de que se tratará todos os preditores como fatores distintos nos permite resu-

mir matrizes de dados em termos de números de (uso) e (não uso de contracepção) em cada um dos dezesseis grupos por combinações de valores da pré-preditores. Para os modelos que envolvem fatores discretos pode-se obter exatamente o mesmo resultados trabalhando com dados agrupados ou com dados individuais, mas o agrupamento é conveniente, pois leva a conjuntos de dados menores.

Tabela 1: A Frequência do uso de contraceptivos entre mulheres casadas por idade, escolaridade e desejo de ter mais filhos (DMF), de Fiji Fertility Survey, 1975.

Idade	Educação	DMF	Y (Uso de Anticoncepcionais)		Total
			Não	sim	
A	B	S	53	6	59
		N	10	4	14
B	A	S	212	52	264
		N	50	10	60
	B	S	60	14	74
		N	19	10	29
C	A	S	155	54	209
		N	65	27	92
	B	S	112	33	145
		N	77	80	157
D	A	S	118	46	164
		N	68	78	146
	B	S	35	6	41
		N	46	48	94
Total	—	S	8	8	16
		N	12	31	43
Total			1100	507	1607

Ajustou-se o primeiro modelo apenas com os efeitos principais (Idade, Escolaridade, DMF) e um segundo modelo com todos os efeitos principais e interações (Idade e Escolaridade, Idade e DMF, Escolaridade e DMF). A metodologia utilizada foi com base nos modelos lineares generalizados, utilizand-se a distribuição binomial e função de ligação logit e probit.

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \quad (3.1)$$

em que,  $\eta$  é meu preditor linear que vai reunir toda informação sobre o uso de anticoncepcional;  $\beta_0$  é o intercepto do modelo;  $\beta_1$  é o coeficiente de regressão para a variável Idade;  $x_1$  é o vetor correspondente aos valores de Idade;  $\beta_2$  é o coeficiente de regressão para a variável Escolaridade;  $x_2$  é o vetor correspondente aos valores

de Escolaridade;  $\beta_3$  é o coeficiente de regressão para a variável DMF;  $x_3$  é o vetor correspondente aos valores de DMF.

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 \quad (3.2)$$

em que, tem todos os efeitos principais de (Idade, Escolaridade e DMF) e suas interações,  $\beta_{12}$  é o coeficiente de regressão para a variável Idade e Escolaridade;  $x_1 x_2$  é o vetor correspondente aos valores de Idade e Escolaridade;  $\beta_{13}$  é o coeficiente de regressão para a variável Idade e DMF;  $x_1 x_3$  é o vetor correspondente aos valores de Idade e DMF;  $\beta_{23}$  é o coeficiente de regressão para a variável Escolaridade e DMF;  $x_2 x_3$  é o vetor correspondente aos valores de Escolaridade e DMF.

### 3.1 Resultados

Os resultados observados nesta seção terão como base os dados de Little (1978). Primeiramente, foi ajustado o modelo (4.1), utilizando a função de ligação logit, de forma que o uso de anticoncepcional dependente da idade, educação e DMF. Os resultados encontram-se na (Tabela 2).

Tabela 2: Interação dos efeitos principais, Idade, Escolaridade e DMF, com ligação Logit.

Variáveis	Preditores	E.P.	Estat. Z	$Pr(>  z )$
Intercepto	-1,9662	0,1720	-11,429	< 0,001
Idade - B	0,3894	0,1759	2,214	0,02681
Idade - C	0,9086	0,1646	5,519	< 0,001
Idade - D	1,1892	0,2144	5,546	< 0,001
Escolaridade - B	0,3250	0,1240	2,620	0,00879
DMF - S	0,8330	0,1175	7,091	< 0,001

A deviance residual encontrada para este modelo foi de 29,92 sobre 10 graus de liberdade. Para checar se o modelo foi bem ajustado procedeu-se um teste de  $\chi^2$ , o valor  $p$  encontrado sobre este quantil foi de  $0,0008 < 0,05$ , e o AIC foi de 113,4. Sendo assim, o modelo aditivo não está bem ajustado aos dados e optou-se por ajustar um modelo com a adição dos termos de interação, entre Idade e Educação, Idade e DMF, e Educação e DMF. A (Tabela 3) apresenta os resultados dos preditores lineares deste novo modelo.

Tabela 3: Interação dos efeitos principais, idade, escolaridade e DMF e suas interações, Idade e Educação, Idade e DMF, e Educação e DMF, com ligação Logit.

Variáveis	Estimativas	E.P.	Estat. Z	Pr(> z )
Intercepto	-1,94437	0,35511	-5,475	< 0,001
Idade - B	0,53041	0,43156	1,229	0,2191
Idade - C	0,62235	0,38983	1,596	0,1104
Idade - D	0,29703	0,49935	0,595	0,5519
Desejo - S	0,47338	0,39674	1,193	0,2328
Educação - S	0,50553	0,37520	1,347	0,1779
Idade - B: DMF - S	0,22536	0,41024	0,549	0,5828
Idade - C: DMF - S	0,95014	0,38199	2,487	0,0129
Idade - D :DMF - S	1,19012	0,51278	2,321	0,0203
Idade - B: Educação - S	-0,15790	0,45838	-0,344	0,7305
Idade - C: Educação - S	-0,05178	0,41840	-0,124	0,9015
Idade - D: Educação - S	0,98645	0,52114	1,893	0,0584
DMF - S: Educação - S	-0,48617	0,26522	-1,833	0,0668

O Desvio do modelo, a deviance residual, foi de 2,45 com 3 graus de liberdade não sendo significativo ao nível de cinco por cento, valor P de 0,48, por isso não se tem nenhuma evidência contra este modelo. O valor de AIC foi de 99,94. Mesmo assim foi realizada a tabela de análise de variância para comparar os modelos. Os resultados se encontram na (Tabela 4).



Tabela 4: Comparação dos modelos aninhados sob a função de ligação logit.

Modelo	GL	Dev. Resid.	Dev. Resid.	Dif. GL	Dif. Dev. Resid.
1		10	29,917		
2		3	2,4415	7	27,476

Por meio da Tabela 4, percebe-se que houve uma redução da deviance residual de 27,476 quando se passa do modelo (4.1) para o modelo (4.2), porém, esta redução teve um custo de 7 graus de liberdade, devido a magnitude da redução na deviance residual, julga-se adequado a inclusão das interações de Idade e DMF, Idade e Educação e Educação e DMF.

Os gráficos de resíduos para ambos os modelos encontram-se nas figuras 1 e 2, respectivamente.

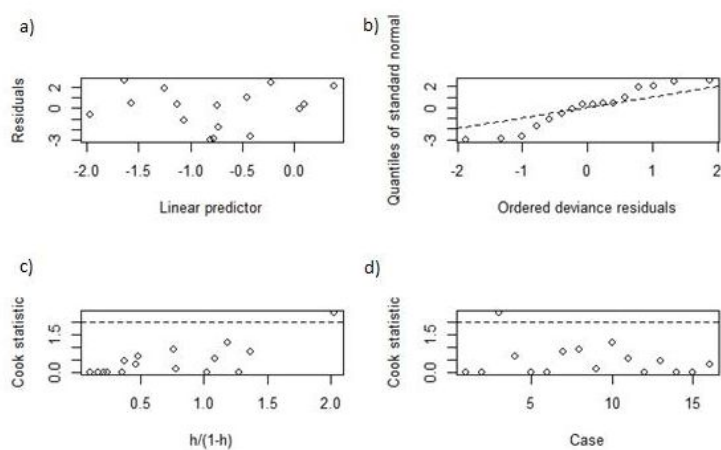


Figura 1: Gráfico de resíduos para o modelo (3.1) com função de ligação logit

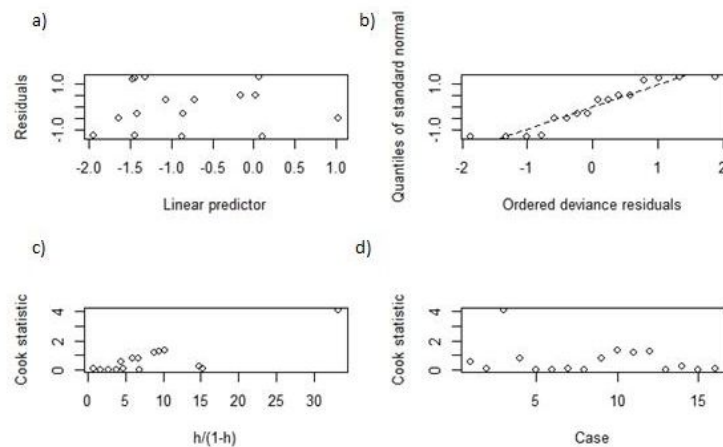


Figura 2: Gráfico de resíduos para o modelo (3.2) com função de ligação logit

Os gráficos de resíduos são similares aos gráficos de resíduos de um modelo linear, porém no modelo linear generalizado os resíduos são obtidos por meio da raiz quadrada da contribuição de uma observação para o desvio (deviance residual). O gráfico dos resíduos na figura 1 (b) percebe-se que o modelo (4.1) não se ajustou bem aos dados pois os pontos ordenados (resíduos) não se ajustaram a reta, um ponto merece maior atenção pois em quase todos os gráficos da figura 1 ele está destacado, havendo indícios de ser ponto de alavanca. Na figura 1(a) percebe-se que há falta de preditores devido a não aleatoriedade dos resíduos com o preditor linear.

Na figura 2 houve uma melhora no ajuste, isto pode ser visto por meio do gráfico dos quantis da normal padrão e os resíduos deviance ordenados (Figura 2(b)), existindo ainda um ponto que se destaca dos demais. Na figura 2(a) percebe-se uma aleatoriedade entre os resíduos e o preditor linear, indicando que a adição dos termos de interação melhorou o ajuste do modelo aos dados.

Na Tabela 5 ao se realizar a análise por meio da função de ligação probit percebe-se que os preditores do modelo (4.1) apresentaram efeito sobre a variável resposta em conformidade com o que foi observado na (Tabela 2). O valor calculado de AIC para este modelo foi de 114,59 valor um pouco maior do que o encontrado

sobre o mesmo modelo com função de ligação logit (Tabela 2). A deviance residual encontrada para este modelo foi de 31,081 sobre 10 graus de liberdade, também foi um pouco maior que a encontrada na função de ligação logit. O valor P encontrado sobre o quantil de 31,081 e 10 graus de liberdade foi  $< 0,001$ , portanto o modelo não está bem ajustado.

Tabela 5: Interação dos efeitos principais, Idade, Escolaridade e DMF, com ligação Probit.

Variáveis	Preditores	E.P.	Estat. Z	$Pr(>  z )$
Intercepto	-1.1786	0,09840	-11,975	$< 0,001$
Idade - B	0,2228	0,10102	2,205	0,0274
Idade - C	0,53049	0,09567	5,545	$< 0,001$
Idade - D	0,70598	0,12819	5,507	$< 0,001$
Escolaridade - B	0,19873	0,07432	2,674	0,0075
DMF - S	0,50284	0,07116	7,066	$< 0,001$

Os resultados encontrados com o modelo (4.2) sob função de ligação probit estão apresentados na (Tabela 6). O valor do AIC encontrado para este modelo foi de 99,81 enquanto que o valor do AIC do modelo (4.1) sob função de ligação probit (Tabela 4) e o valor do AIC do modelo (4.2) sob função de ligação logit (Tabela 3) foram maiores, indicando que este modelo (4.2) sobre função de ligação probit foi o mais adequado dentre todos os modelos avaliados. Percebe-se também que ao se comparar os resultados encontrado pelo modelo (4.1) em ambas as funções de ligação (logit e probit), a primeira apresenta valor menor, fato que não se verifica na mesma comparação do modelo (4.2), dando indícios que ao se parametrizar mais o modelo a função de ligação probit se adequa melhor a estes dados.

Tabela 6: Interação dos efeitos principais, idade, escolaridade e DMF e suas interações, Idade e Educação, Idade e DMF, e Educação e DMF, com ligação Probit.

Variáveis	Estimativas	E.P.	Estat. Z	Pr(> z )
Intercepto	-1,15960	0,19171	-6,049	< 0,001
Idade - B	0,30048	0,23880	1,258	0,20829
Idade - C	0,35746	0,21398	1,670	0,09482
Idade - D	0,16201	0,27758	0,584	0,55945
Desejo - S	0,27883	0,22776	1,224	0,22087
Educação - S	0,28847	0,20438	1,411	0,15813
Idade - B: DMF - S	0,13638	0,23686	0,576	0,56477
Idade - c: DMF - S	0,58844	0,22013	2,673	0,00752
Idade - D :DMF - S	0,72941	0,29755	2,451	0,01423
Idade - B: Educação - S	-0,08451	0,25630	-0,330	0,74160
Idade - C: Educação - S	-0,02338	0,23313	-0,100	0,92010
Idade - D: Educação - S	0,61503	0,29919	2,056	0,03982
DMF - S: Educação - S	-0,28863	0,15841	-1,822	0,06845

As figuras 3 e 4 apresentam os gráficos dos resíduos deviance dos modelos (4.1) e (4.2) sob a função de ligação probit. Percebe-se que similar ao que ocorreu com os resíduos deviance dos modelos sob função de ligação logit o modelo 4.2 ficou melhor ajustado.

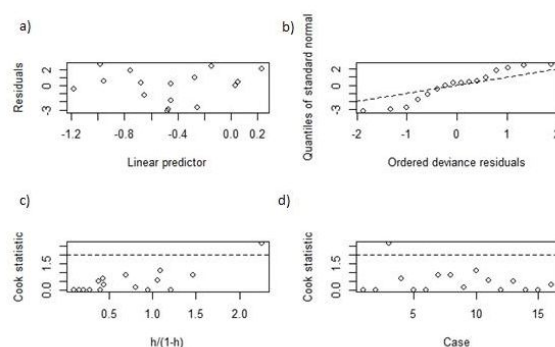


Figura 3: Gráfico de resíduos para o modelo (3.1) sob função de ligação probit

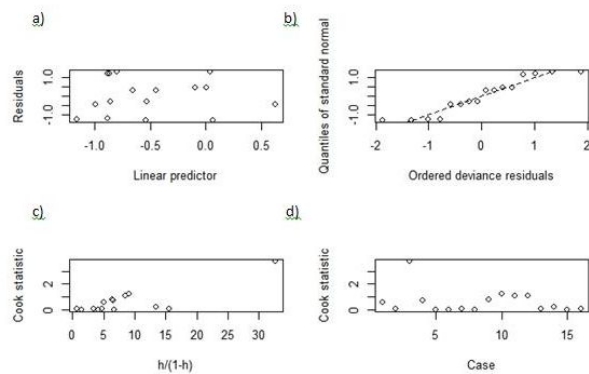


Figura 4: Gráfico de resíduos para o modelo (3.2) sob função de ligação probit

Entre as mulheres que desejam ter filhos e pertencem as faixas de idade (C e D), pelo modelo selecionado (4.2) sob função de ligação probit (Tabela 5), há uma diminuição média de 12% ( $\Phi^{-1}(0,88)$ ) no uso de anticoncepcionais.

## 4 Conclusão

Embora seja bastante comum encontrarmos conjuntos de dados que apresentem distribuições diferentes da normal, e que uma quantidade razoável de material já tenha sido publicada nessa área, a análise de diagnóstico é pouca explorada. O enfoque computacional nesse caso é de uma enorme importância, dada a complexidade dos algoritmos a serem utilizados. Um segundo objetivo foi o de ajustar estes modelos com um algoritmo implementado no software livre R. A linguagem R é bastante intuitiva e versátil, mostrou-se muito eficaz com as funcionalidades já desenvolvidas para os modelos aqui estudados (Normal, Poisson, Binomial Negativa, Normal Inversa e Gama) facilitando de certa maneira o trabalho desenvolvido.

Para a característica binária utilizada neste estudo o modelo linear generalizado mostrou-se adequado para o ajuste dos dados. O modelo que contempla os efeitos principais mais interações duas a duas foi o melhor modelo em relação aos demais na função de ligação probit.

## Referências

- ARANDA-ORDAZ, F. *On the families of transformations to additivity for binary response data*. *Biometrika*, 68, 357-363, 1981.
- ASHTON, W. D. *The Logit Transformation with Special Reference to its Uses in Bioassay*. Gritn, London. 1972.
- BERKSON, J. *Application of the logistic function to bioassay*. *J. R. Statist. Soc. B*, 39, 357-365, 1944.
- BERNOULLI, J. *Ars conjectandi*. Thurnisius, Basilea. 1713.
- BIRCH, M. W. *Maximum likelihood in three-way contingency tables*. *J. R. Statist. Soc. B*, 25, 220-233, 1963.
- BLISS, C. I. *The calculator of the dosage-mortality curve*. *Ann. Appl. Biol.*, 22, 134-167, 1935.
- BOLFARINE, H.; SANDOVAL, M. C. *Introdução à Inferência Estatística*, Coleção Matemática Aplicada – Sociedade Brasileira de Matemática, 2001.
- BUSE, A. *The likelihood ratio, Wald and Lagrange multiplier tests: an expository note*. *The American Statistician* 36, 153-157, 1982.
- CASELLA, G.; BERGER, R. L. *Inferência estatística - tradução da 2ª edição norteamericana*. Centage Learning, 612 p,2011.
- CORDEIRO, G. M.; DEMÉTRIO, C. G. B. *Modelos lineares generalizados*. Minicurso para o 12º SEAGRO e a 52ª Reunião anual da RBRAS UFSM, Santa Maria, RS. p. 1-46. 2010.

- CORDEIRO, G. M.; CRIBARI-NETO, F.; AUBIN, E. Q.; FERRARI, S. L. P. *Bartlett corrections for oneparameter exponential family models*. Journal of Statistical Computation and Simulation, 53, 211-231, 1995.
- CORDEIRO, G.M. *Modelos Lineares Generalizados*. Campinas, VII SINAPE, 286p. 1986.
- CORDEIRO, G. M. *Improved likelihood ratio statistics for generalized linear models*. J. Roy. Statist. Soc. B, 45, 401-413, 1983.
- CORDEIRO, G.M.; LIMA NETO, E.A.. *Modelos Paramétricos*. Recife: Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, 2006.
- CORDEIRO, G. M.; MCCULLAGH, P. *Bias correction in generalized linear models*. J. Roy. Statist. Soc. B, 53, 629-643, 1991.
- COX, D. R. *Analysis of binary data*. Chapman and Hall, London. 1970.
- COX, D. R. *Regression models and life tables (with discussion)*. J. R. Statist. Soc. B, 74, 187-220, 1972.
- COX, D.R.; HINKLEY, D.V. *Theoretical Statistics*. Chapman and Hall, London. 1974.
- DEMÉTRIO, C. G. B. *Modelos lineares generalizados em experimentação agrônômica*. ESALQ/USP - Piracicaba, SP. p.1-54. 2002.
- DOBSON, A. J. *An Introduction to Statistical Modelling*. London: Chapman and Hall, 2<sup>nd</sup> ed., 174p. 1990.
- DYKE, G.; PATTERSON, H. *Analysis of factorial arrangements when the data are proportions*. Biometrics, 8, 1-12, 1952.
- FEIQL, P.; ZELEN, M. *Estimation of exponential survival probabilities with concomitant information*. Biometrics, 21, 826-838, 1965.
- FISHER, R. *On the mathematical foundations of theoretical statistics*. Philo-



- sophical Transactions of the Royal Society, 222, 309-368, 1922.
- FISHER, R. *The case of zero survivors (appendix to bliss, c.i.)*. Ann. Appl. Biol., 22, 164-165, 1935.
- GASSER, M. *Exponential survival with covariance*. Journal of the American Statistical Association, 62, 561-568, 1967.
- HABERMAN, S. *The general log-linear model*. PhD dissertation. Univ. of Chicago Press, Chicago, Illinois. 1970.
- JORGENSEN, B. *Maximum likelihood estimates and large samples inference for generalized linear and nonlinear regression models*. Biometrika, 70, 19-28, 1987.
- LINDSEY, J.K. *Applying Generalized Linear Models*. Springer, New York. 1997.
- LITTLE, R. J. A. *Generalized Linear Models for Cross-Classified Data from the WFS*. World Fertility Survey Technical Bulletins, 5, 1978.
- MCCULLOCH, C. E.; SEARLE, S.R. *Generalized, Linear, and Mixed Models*. New York: John Wiley e Sons, 325p. 2001.
- MCCULLAQH, P.; NELDER, J. A. *Generalized Linear Models*. Chapman and Hall, London, second edition. 1989.
- MENDENHALL, P.; SCHEA®ER, R. L.; WACKERLY, D. D. *Mathematical Statistics with Applications*. Duxbury, Boston. 1981.
- MORRIS, C. N. *Natural exponential families with quadratic variance functions: statistical theory*. Annals of Statistics, 11, 515-529. 1982.
- NELDER, J. A.; WEDBERBURN, R. W. M. *Generalized linear models*. Journal of the Royal Statistical Society, A, 135, 370-384, 1972.
- PAULA, G. A. *Modelos de Regressão com Apoio Computacional*. São Paulo: IME/USP. 2000.
- PAULA, G.A.. *Modelos de Regressão com Apoio Computacional, São Paulo:*

IME – Universidade de São Paulo, 2010.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, 2013. Disponível em: <<http://www.Rproject.org/>>. Acesso em 1 de Janeiro de 2014.

RASCH, G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Danmarks Paedogogiske Institut, Copenhagen. 1960.

SILVEY, S. D. *Statistical Inference*. London: Chapman and Hall, 2<sup>nd</sup> ed., 191p. 1975.

WALD, A. *Tests os statistical hypotheses concerning several parameters when the number of observations is large*. *Trans. Amer. Math. Soc.*, 54, p.426-482, 1943.

ZIPPIN, C.; ARMITAGE, P. *Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter*. *Biometrics*, 22, 665-672, 1966.

# APÊNDICE

**Comandos utilizados para a análise do banco de dados cuse.txt.**

```
cuse <- read.table("http://data.princeton.edu/wws509/datasets/cuse.dat",
header=TRUE)

# Carregar biblioteca e ler os dados #

library(MASS)
cuse
attach(cuse)

noMore <- wantsMore == "no"
hiEduc <- education == "high"

# Análise modelo Logit #

lrfit0=glm( cbind(using,notUsing)  age + hiEduc + noMore, family=binomial)
summary(lrfit0)
1-pchisq(29.92,10)

lrfit0 <- glm( cbind(using,notUsing)  age * noMore + hiEduc + age:hiEduc+noMore:hiEduc,
family=binomial) summary(lrfit0)
1-pchisq(2.44,3)
summary(lrfit0)

# Análise modelo Probit #

lrfitp <- glm( cbind(using,notUsing)  age + hiEduc + noMore , family=binomial(link=probit))
summary(lrfitp)
1-pchisq(12.63,7)
```

```
lrfit0p <- glm( cbind(using,notUsing) ~ age * noMore + hiEduc + age:hiEduc+noMore:hiEduc,  
family=binomial(link=probit))  
summary(lrfit0p)  
1-pchisq(2.44,3)  
summary(lrfit0p)  
  
par(mfrow=c(2,2))  
plot(lrfit0p)
```