Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Modelos para dados de proporção: um estudo sobre a viabilidade de ovos do mosquito *Aedes aegypti*, mantidos em diferentes tipos de armazenamentos na Amazônia

Pedro Marinho Amoêdo

Dissertação apresentada para obtenção do título de Mestre em Ciências. Área de concentração: Estatística e Experimentação Agronômica

Piracicaba
2014

Pedro Marinho Amoêdo
Bacharel em Estatística

Modelos para dados de proporção: um estudo sobre a viabilidade de ovos do mosquito *Aedes aegypti*, mantidos em diferentes tipos de armazenamentos na Amazônia

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientadora:
Prof^ª Dr^a **SÔNIA MARIA DE STEFANO PIEDADE**

Dissertação apresentada para obtenção do título de Mestre em Ciências. Área de concentração: Estatística e Experimentação Agronômica

**Piracicaba
2014**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - ESALQ/USP**

Amoêdo, Pedro Marinho

Modelos para dados de proporção: um estudo sobre a viabilidade de ovos do mosquito *Aedes aegypti*, mantidos em diferentes tipos de armazenamentos na Amazônia / Pedro Marinho Amoêdo. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2014.
78 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2014.
Bibliografia.

1. MLG binomial 2. Dados na forma proporção 3. Envelope de simulação 4. Ovos de *Aedes aegypti* I. Título

CDD 614.571
A523m

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

*Dedico a meus pais,
Rossi Paes de Andrade Amoêdo e
Iolanda Marinho Amoêdo.*

AGRADECIMENTOS

A Deus e a minha grande mãe Maria por sempre estarem presentes em cada momento de minha vida, mesmo quando eu não os via, sei que estavam lá.

Aos meus pais, por todos os ensinamentos e por tudo a mim dedicado. Sei que muitas vezes deixaram de realizar seus desejos em detrimento dos meus.

À minha família como um todo, que de forma muito especial me incentivou com muito amor e carinho, dando apoio e suporte a cada dia.

À Bia, minha eterna indiazinha. Amiga e companheira que sempre esteve ao meu lado, meu porto seguro.

À professora Dra. Taciana Villela Savian que me conduziu sabiamente nesta minha impleitada com muito carinho e respeito.

Aos professores do LCE/ESALQ/USP, que estiveram presente neste tempo de curso. Em especial a professora Dra. Sônia Maria de Stefano Piedade, não só pela sua orientação, mas sobretudo pela professora e pessoa que é.

Aos amigos de Manaus e Parintins que mesmo na distância sempre me apoiaram, em particular aos amigos Lúcio Pontes e Wallace Mendes.

Aos colegas de departamento que sempre incentivaram a caminhada. Em especial aos amigos: Djair Durand, Edilan Quaresma, Elias Medeiros, Gabriel Avancini, José Nilton, Kuang Hongyu, Marcos Maverick, Marcello Neiva, Rafael Moral, Reginaldo Hilário, Ricardo Klein, Thiago Oliveira, Simone Grego, Maria Cristina. Estes vivenciaram comigo problemas, estudos e acima de tudo me ajudaram com suas partilhas.

À FAPEAM - Fundação de Amparo à Pesquisa do Estado do Amazonas, pela bolsa de mestrado concedida.

A todos os professores e profissionais da educação que estiveram presente em minha vida, em particular a professor Dra. Rosana Cristina Pereira Parente da Universidade Federal do Amazonas, que com respeito, carinho e dedicação a profissão me incentivou continuar os estudos.

E a todos que contribuíram de alguma forma para a minha formação.

O meu muito obrigado!

*“Se a riqueza é um bem desejável na vida,
que há de mais rico que a sabedoria que tudo criou?
Se a inteligência do homem consegue operar,
o que, então, mais que a sabedoria, é artífice dos seres?
E se alguém ama a justiça,
seus trabalhos são virtudes;
ela ensina a temperança e a prudência,
a justiça e a força:
não há ninguém que seja mais útil aos homens na vida.
Se alguém deseja uma vasta ciência,
ela sabe o passado e conjectura o futuro;
conhece as sutilezas oratórias e resolve os enigmas;
prevê os sinais e os prodígios,
e o que tem que acontecer no decurso das idades e dos tempos.
Portanto, resolvi tomá-la por companheira de minha vida,
cuidando que ela será para mim uma boa conselheira.”*

Sabedoria 8,5-9

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	17
1 INTRODUÇÃO	19
2 REVISÃO BIBLIOGRÁFICA	21
2.1 <i>Aedes aegypti</i>	21
2.2 Modelo linear generalizado	23
2.2.1 Estrutura do modelo	24
2.3 Família exponencial	25
2.4 Método de Estimação	26
2.4.1 Método de Newton-Rapshon	27
2.4.2 Método escore de Fisher	28
2.4.3 Critério de convergência	30
2.5 Medidas da qualidade do ajuste	30
2.5.1 Análise de desvio	32
2.6 Análise de resíduos e diagnósticos	34
2.7 Adequação da função de ligação	37
2.8 Adequação da função de variância	37
3 MATERIAL E MÉTODOS	39
3.1 Material	39
3.1.1 Testes de viabilidade	39
3.2 O experimento	41
3.3 Métodos	42
3.4 Modelo linear generalizado binomial	42
3.5 Modelos ajustados	43
3.6 Modelo binomial	43
3.7 Equações de verossimilhança do MLG binomial	44
3.8 Ajuste dos modelos	46
3.9 Desvio do MLG binomial	46
3.10 Componentes do desvio do MLG binomial	47

3.11 Análise de desvio - ANODEV	48
3.12 Viabilidade dos ovos de <i>Aedes aegypti</i>	48
3.13 Intervalo de confiança para DL_{50}	49
3.13.1 Método Delta	49
3.14 Casos especiais do MLG binomial	50
3.14.1 Modelo logístico	50
3.14.2 Modelo probito	51
4 RESULTADOS E DISCUSSÃO	53
4.1 Modelos ajustados	54
5 CONCLUSÕES	65
REFERÊNCIAS	67
ANEXOS	71

RESUMO

Modelos para dados de proporção: um estudo sobre a viabilidade de ovos do mosquito *Aedes Aegypti*, mantidos em diferentes tipos de armazenamentos na Amazônia

O presente trabalho teve por objetivo estimar o tempo que os ovos de *Aedes aegypti* permanecem em condições de eclodir e produzir larvas viáveis após permanecerem armazenados por determinado tempo (viabilidade) em diferentes tipos de recipientes. Para tal, usou-se a metodologia dos modelos lineares generalizados para dados na forma de proporção. Como função de ligação, optou-se pelas funções logística e probito, como preditores lineares retas paralelas e retas concorrentes. Foi realizado um experimento, em que procurou-se simular, tanto quanto possível, o efeito das condições ambientais que os ovos de *Aedes aegypti* são submetidos. Os ovos foram armazenados em três recipientes (copo plástico, saco plástico e envelope de papel) e tratados com água ao final de cada período de armazenamento estabelecido. Por meio do ajuste do modelo linear generalizado binomial com função de ligação logística e preditor linear retas paralelas, foi possível prever a viabilidade dos ovos de *Aedes aegypti*. Como medidas da qualidade do ajuste, usou-se o desvio residual com valor de p (nível descritivo), calculado sobre os percentis da distribuição qui-quadrado e o gráfico de probabilidade normal padrão com envelope de simulação. Verificou-se comportamento semelhante entre saco plástico e envelope de papel com respeito a viabilidade estimada dos ovos de *Aedes aegypti*, ambos diferindo do recipiente copo plástico que apresentou a maior viabilidade.

Palavras-chave: MLG binomial; Dados na forma proporção; Envelope de simulação; Ovos de *Aedes aegypti*

ABSTRACT

Models for proportion data: a study about the viability of *Aedes aegypti* mosquito eggs, kept in different storage types in the Amazon

This work had like objective estimate the time in which the *Aedes aegypti* eggs remain in hatch conditions and produce viable larvae after remain a certain time (viability) in different storage types. To this, the generalized linear models methodology for data in proportion form was used. As link function was chose the logistic and probit and like linear predictors, parallel and intersecting lines. Was conducted an experiment, in which were simulated as much as possible the effect of environmental conditions in which the *Aedes aegypti* eggs are submitted. The eggs were stored in three containers (plastic cup, plastic bag and paper envelope) and treated with water at the end of each storage period. With the fit of the binomial generalized linear model with logistic link function and parallel lines like linear predictor, was possible predict the viability of the *Aedes aegypti* eggs. As quality measures of fit, the residual deviance with p-value (descriptive level), calculated on the chi-square distribution percentiles and the standard normal probability plot with simulated envelope were used. It has been found similar behaviour between plastic bag and paper envelope with respect to estimated viability of the *Aedes aegypti* eggs, both differing from the plastic cup that presents the highest viability.

Keywords: Binomial GLM; Data in proportion form; Simulation envelope; *Aedes aegypti* eggs

LISTA DE FIGURAS

Figura 1 - Tipos de recipientes em que os ovos de <i>Aedes aegypti</i> foram armazenados: envelopes de papel, copos plástico e sacos plástico.	40
Figura 2 - Local rodeada por vegetação e com cobertura em que os lotes dos ovos de <i>Aedes aegypti</i> foram mantidos por diferentes períodos de armazenamento	40
Figura 3 - Gráfico de dispersão dos dados, proporções de eclosões de larvas <i>versus</i> idade dos ovos (período que receberam o tratamento)	53
Figura 4 - Gráficos das proporções de larvas de <i>Aedes aegypti versus</i> a idade em que os ovos foram tratados e as curvas ajustadas pelo MLG com ligação probita, utilizando-se os preditores lineares retas concorrentes (a) e retas paralelas (b)	54
Figura 5 - Gráficos das proporções de larvas do <i>Aedes aegypti versus</i> a idade em que os ovos foram tratados e as curvas ajustadas pelo MLG com ligação logística, utilizando-se os preditores lineares retas concorrentes (a) e re- tas paralelas (b)	55
Figura 6 - Gráficos de probabilidade normal com envelopes simulados referentes ao MLG binomial com função de ligação probita e preditores lineares retas concorrentes (a) e retas paralelas (b)	56
Figura 7 - Gráficos das proporções de larvas de <i>Aedes aegypti versus</i> a idade em que os ovos foram tratados e as curvas ajustadas pelo MLG com in- clusão de peso na variável binomial com ligação probita e preditores retas concorrentes (a) e retas paralelas (b)	56
Figura 8 - Gráficos de probabilidade normal com envelopes simulados referentes ao MLG binomial com função de ligação logística e preditores lineares retas concorrentes (a) e retas paralelas (b)	58
Figura 9 - Gráficos das proporções de larvas de <i>Aedes aegypti versus</i> a idade em que os ovos foram tratados e as curvas ajustadas, considerando-se a ligação logística e os preditores lineares retas concorrentes (a) e retas paralelas (b)	58
Figura 10 - Gráficos de probabilidade normal com envelopes simulados referentes ao MLG binomial com função de ligação logística e preditores lineares retas concorrentes (a) e retas paralelas (b)	60

Figura 11 - Gráficos das proporções de larvas do *Aedes aegypti* versus a idade em que os ovos foram tratados e as curvas ajustadas pelo MLG com ligação logística, utilizando-se como preditores lineares retas concorrentes (a) e retas paralelas (b) 61

LISTA DE TABELAS

Tabela 1 - Construção da análise de desvio - ANODEV, para modelos lineares generalizados	33
Tabela 2 - Frequência de larvas e total dos ovos de <i>Aedes aegypti</i> produzidos em condições laboratoriais, segundo tipo de recipientes e período de armazenamento	41
Tabela 3 - Análise de desvio para o MLG binomial com função de ligação probito e preditores lineares retas paralelas	57
Tabela 4 - Análise de desvio para o MLG binomial com função de ligação probito e preditores lineares retas concorrentes	57
Tabela 5 - Análise de desvio para o MLG binomial com função de ligação logística e preditores lineares retas paralelas	59
Tabela 6 - Análise de desvio para o MLG binomial com função de ligação logística e preditores lineares retas concorrentes	59
Tabela 7 - Desvios residuais do MLG binomial com função de ligação logística . . .	61
Tabela 8 - Análise de desvio para o MLG binomial com função de ligação logística e preditor linear retas concorrentes	61
Tabela 9 - Análise de desvio para o MLG binomial com função de ligação logística e preditor linear retas paralelas	62
Tabela 10 - Estimativas dos parâmetros do MLG binomial com função de ligação logística e preditor linear retas paralelas considerando todas as observações	62
Tabela 11 - Estimativas dos parâmetros do MLG binomial com função de ligação logística e preditor linear retas paralelas em que eliminou-se duas observações	63
Tabela 12 - Análise de desvio para o novo modelo em que utilizou-se a função de ligação logística e preditor linear retas paralelas	64

1 INTRODUÇÃO

A dengue é uma doença viral dos países tropicais e subtropicais de evolução geralmente benigna. É transmitida ao homem pela picada da fêmea do mosquito *Aedes aegypti*, vetor dos quatro sorotipos do vírus da dengue. A maneira usual de combate a dengue na atualidade é por meio do controle das densidades populacionais do mosquito *Aedes aegypti* mediante aplicação de produtos químicos. No combate das larvas aplica-se diretamente nos criadouros potenciais e para o combate dos alados, diretamente no ar, pois infelizmente ainda não existe uma vacina de prevenção para a doença.

O estudo das características do *Aedes aegypti*, bem como seus padrões de comportamento e desenvolvimento nas diferentes fases do seu ciclo de vida, constitui ferramenta importante para a compreensão da dinâmica populacional desta espécie. Sabendo-se que a fase de ovo representa a de maior resistência do ciclo biológico do mosquito e o principal fator que favorece a dispersão mundial do vetor, tornam-se necessários estudos relativos à viabilidade dos ovos nas condições ambientais, a fim de se obterem informações que possam melhorar o direcionamento das ações de controle.

Na busca dessas informações, realizou-se testes com ovos do mosquito *Aedes aegypti* no laboratório de malária e dengue em Manaus(AM) com propósito de medir a viabilidade destes sob diferentes formas de armazenamento: ovos estocados em copos, em envelopes e sacos plásticos. No estudo, tomou-se como variável resposta para análise estatística a proporção de larvas provenientes dos totais de ovos armazenados.

Um caso particular de análise de dados é aquele em que a variável resposta é uma proporção, e podendo esta assumir naturalmente distribuição binomial. Dentre as várias técnicas estatísticas que podem ser aplicadas para a análise, tem-se a classe dos modelos lineares generalizados, que é a unificação de uma ampla variedade de métodos estatísticos proposta por (NELDER : WEDDERBURN, 1972). como regressão, análise de variância. Esta classe considera tanto modelos com variáveis respostas categóricas como numérica; considera distribuições como binomial, Poisson, normal, entre outras.

Muitas vezes, a formulação de um modelo linear generalizado (MLG) para experimentos contendo dados na forma de proporção não é simples, pois existe a possibilidade da ocorrência de uma variabilidade maior do que a esperada pelo modelo, em casos como este, a solução mais comum é de se estimar um modelo que melhor acomode essa variação extra. Ainda, pode ocorrer que a simples eliminação de pontos aberrantes

resolva o problema (PAULA, 2013). Um modelo linear generalizado é formulado a partir de três componentes: componente aleatória, componente sistemática e uma função que relaciona a componente aleatória à componente sistemática, denominada de função de ligação, a qual é escolhida levando-se em consideração a natureza da variável Y ou a natureza da modelagem a ser utilizada.

Os parâmetros do modelo linear generalizado podem ser estimados por máxima verossimilhança via métodos iterativos como o de Newton-Raphson ou escore de Fisher. Este último, também é denominado de processo iterativo de mínimos quadrados ponderados. Este processo iterativo é sem dúvida bem atrativo e preferido, dada a facilidade de estimação dos parâmetros por rotinas contidas em pacotes estatísticos, além do que, o processo iterativo é válido para qualquer MLG. O método de máxima verossimilhança também estima a matriz de covariância assintótica do modelo.

Neste trabalho, é apresentado o ajuste de modelos lineares generalizados, com vista à produção de informações que possam auxiliar no direcionamento das ações de controle do mosquito *Aedes aegypti*. Para verificar o ajuste dos modelos, foram usados os desvios residuais e os gráficos de probabilidade normal padrão com envelopes de simulações e como medida de comparação/seleção a análise de desvio, uma generalização da análise de variância para MLGs.

Este trabalho teve como objetivos estimar o tempo que os ovos de *Aedes aegypti* permanecem em condições de eclodir e produzir larvas viáveis após permanecerem armazenados por determinado tempo (viabilidade) e apresentar a metodologia dos modelos lineares generalizados para análise de dados na forma de proporções, com respeito ao ajuste, medida da qualidade de ajuste, análise de desvio e análise confirmatória do ajuste.

Na segunda seção é feita uma revisão da literatura à respeito de algumas características do vetor *Aedes aegypti* e modelos lineares generalizados, na seção 3 é descrito o experimento, o modelo linear generalizado binomial, ajustes, seleção de modelos e qualidade do ajuste. Na seção 4 são realizadas as análises do conjunto de dados e na seção 5 são apresentadas as conclusões a respeito do trabalho.

2 REVISÃO BIBLIOGRÁFICA

2.1 *Aedes aegypti*

Dentre as doenças chamadas reincidentes, a dengue configura a mais importante arbovirose que afeta o homem e constitui sério problema de saúde pública. Particularmente nos países tropicais, onde as condições ambientais, associadas à ineficácia das políticas públicas de saúde, favorecem o desenvolvimento e a proliferação do *Aedes aegypti*, principal vetor (SAMPAIO, 2010).

Esta doença se manifesta como uma doença febril aguda de evolução benigna na forma clássica, e grave, quando se apresenta na forma hemorrágica. Tem como agente etiológico um arbovírus do gênero Flavivírus da família Flaviviridae, do qual são conhecidos quatro sorotipos do vírus da dengue: DEN-1, DEN-2, DEN-3 e DEN-4. A infecção por um deles confere proteção permanente para o mesmo sorotipo e imunidade parcial e temporária contra os outros três. Trata-se, caracteristicamente, de enfermidade de áreas tropicais e subtropicais, onde as condições do ambiente favorecem o desenvolvimento dos vetores (Fundação Nacional de Saúde - FUNASA, 2001).

A dengue é transmitida ao homem por mosquitos do gênero *Aedes*, sendo o *Aedes aegypti* seu principal vetor. O *Aedes aegypti* é encontrado, principalmente no meio urbano, vive preferencialmente dentro das casas, colonizado em depósitos de armazenamento de água e pequenas coleções temporárias (latas, pneus, cacos de vidro e vasos de plantas) (BRAGA; VALLE, 2007). Embora, o *Aedes aegypti* seja um mosquito urbano, já foi encontrado em zonas rurais. Acredita-se que para lá tenha sido levado em recipientes ainda na forma de ovos ou larvas.

A transmissão da dengue ao homem se faz pela fêmea do *Aedes aegypti*, durante a hematofagia. A alimentação sanguínea é necessária para completar o processo de amadurecimento do folículo ovariano. De oito a 12 dias após um repasto de sangue infectado, o mosquito está apto a transmitir o vírus da dengue (MONATH, 1994).

O *Aedes aegypti* possui um curto ciclo biológico com duração de 15 a 30 dias, em regiões tropicais que compreende as fases de ovo, quatro estágios larvais (L1, L2, L3 e L4), pupa e adulto (SANTOS, 2008). Ainda, apresenta um ciclo aquático que é influenciado pelo tipo e qualidade dos reservatórios de água. Este inseto prefere reproduzir em reservatórios de águas limpas, embora possa se adaptar às novas situações impostas

pelo homem.

O ovo de *Aedes aegypti*, mede aproximadamente 1mm de comprimento e é depositado um a um pela fêmea em água parada, estes ficam aderidos na superfície interna da parede dos criadouros, preferencialmente em superfícies rugosas, acima da linha da água. Em condições normais, os ovos se desenvolvem, amadurecem e, logo após a imersão na água, eclodem. Ao fim do desenvolvimento, se ocorrer situações adversas, como: dessecação, baixas temperaturas e insolação, o embrião entra no estado da diapausa, ou seja, o adiamento da eclosão. Em condições favoráveis de temperatura e umidade, esse estado de quiescência poderá se prolongar por seis meses ou mais tempo, até que ocorra o contato com a água e eclosão do ovo (FORATTINI, 2002).

Essa característica, que é um dos principais obstáculos para o seu controle, proporcionou a disseminação desse mosquito por amplas áreas geográficas, por meio do transporte de ovos em vários tipos de materiais (pneus, vasos, garrafas e dentre outros). Essa é a chamada dispersão passiva, que pode acontecer por qualquer meio de transporte aéreo, marítimo ou terrestre. Provavelmente foi dessa forma que ocorreu a introdução do *Aedes aegypti* no Brasil, por meio dos navios que transportavam os escravos provenientes do continente Africano nos séculos XVI e XIX (FORATTINI, 2002).

Nos primeiros estudos sobre a oviposição de *Aedes aegypti*, conseguiu-se mostrar a eclosão de ovos mantidos secos por período de dois meses. Seguiram-se outras séries de estudos em que foi constatada que a viabilidade dos ovos é bem maior: observou-se eclosão de ovos mantidos secos por período de até 262 dias (CHRISTOPHERS, 1960 apud PINHEIRO, 2005)

No Brasil, são restritos os trabalhos enfocando períodos de viabilidade. Os mais recentes foram realizados por Silva et al., (1993) e Silva e Silva (1999) que estudaram a influência da dessecação sobre a viabilidade, o período de quiescência dos ovos e outros aspectos relacionados ao ciclo evolutivo do mosquito. Os resultados mostram que 85% deles podem se romper após três dias de quiescência, quando em contato com a água, e menos de 10% podem sobreviver por até 492 dias. A eclosão das larvas, segundo os autores, ocorrem ao longo de aproximadamente 10 dias.

Pinheiro (2005), avaliou a viabilidade dos ovos de *Aedes aegypti* mantidos em áreas interna e externa, em diferentes tipos de períodos e armazenamento. Os resultados mostram comportamento semelhante para a viabilidade ocorrida nos recipientes

saco plástico e envelope de papel diferindo do recipiente copo, que apresentou a maior viabilidade média. Ainda, apresenta um período médio de viabilidade de 98 dias, não fazendo distinção se a área é interna ou externa.

Infelizmente ainda não existe vacina para a prevenção de dengue. Assim, o combate ao *Aedes aegypti* continua sendo a única forma de se prevenir o avanço da doença por meio da eliminação dos criadouros potenciais dos mosquitos vetores, por aplicação de larvicidas em depósitos de água de consumo, uso de inseticidas para as formas adultas durante os períodos de transmissão e um processo contínuo e sustentado de educação das comunidades (TAUIL, 2001).

2.2 Modelo linear generalizado

Nelder e Wedderburn (1972) propuseram uma teoria unificadora da modelagem estatística a qual deram o nome de Modelos Lineares Generalizados (MLGs), como uma extensão dos modelos lineares clássicos. Na realidade, eles mostraram que uma série de técnicas comumente estudadas separadamente podem ser reunidas sob o nome de modelos lineares generalizados. Mostraram ainda, que grande parte dos problemas estatísticos que surgem em diversas áreas da pesquisa podem ser formulados de uma única maneira. Além de Nelder e Wedderburn (1972), na literatura, existem diversos livros clássicos que tratam de MLG, como exemplo, (AGRESTI, 2002; COLLETT, 2003; DOBSON, 2002; McCULLAGH ; NELDER, 1989) e em língua portuguesa pode-se encontrar (CORDEIRO; DEMÉTRIO, 2011; PAULA, 2013).

Os modelos lineares generalizados fornecem uma estrutura teórica geral para muitos modelos estatísticos, permitem o ajuste de modelos de regressão para uma variável resposta que pode assumir diferentes tipos de distribuições, além do que, também, simplificam a implementação desses modelos em diferentes *softwares* estatísticos, uma vez que, os mesmos algoritmos podem ser utilizados (JACKMAN, 2011).

Segundo Myers e Montgomery (1997), os modelos lineares generalizados são usados quando os erros não seguem uma distribuição normal e a suposição de homogeneidade for violada. Esses modelos envolvem uma variável resposta e um conjunto de variáveis explicativas, sendo que a variável resposta, componente aleatória do modelo, pertence à família exponencial, e as variáveis explicativas entram na forma de uma função linear (componente sistemática) relacionada à componente aleatória a partir de

uma função denominada de função de ligação.

Usualmente, quando há violação da suposição de homogeneidade de variância, faz-se transformação da variável resposta de modo a melhorar o ajuste (BOX; COX, 1964). Modelos baseados em transformações podem apresentar problemas nos valores estimados e nos intervalos de confiança (MYERS; MONTGOMERY, 1997). Ainda, salientam que nos MLGs, os intervalos de confiança são uniformemente menores, trazem mais informações, sugerindo um modelo mais eficiente. Vieira (2004) enfatiza que os MLGs são mais flexíveis, no sentido de considerar outras distribuições que não somente a distribuição normal, não exige variância constante, pode ser qualquer função da média.

2.2.1 Estrutura do modelo

Segundo Agresti (2002), Paulino e Singer (2006), entre outros um modelo linear generalizado é especificado por três componentes:

1. Componente aleatória: consiste de um conjunto de variáveis aleatórias independentes $Y_1, Y_2, \dots, Y_n$ (mas não identicamente distribuídas) obtidas de uma mesma distribuição que pertence à família exponencial de distribuições, com médias $\mu_1, \mu_2, \dots, \mu_n$, ou seja,

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

$$i = 1, 2, \dots, n;$$

2. Componente sistemática: a componente sistemática do modelo específica a estrutura linear das variáveis explicativas (quantitativas e/ou qualitativas), as quais entram no modelo na forma de uma soma linear de seus efeitos, dando origem ao preditor linear (McCULLAGH ; NELDER, 1989)

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

$i = 1, 2, \dots, n$ ou, em termos matriciais, por

$$\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^T = \mathbf{X} \boldsymbol{\beta},$$

em que $\mathbf{X}$ é a matriz cujas linhas $\mathbf{x}_i^T$ são os valores das variáveis explicativas para a estrutura paramétrica da distribuição de todos os $\{y_i\}$, e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ é o vetor de parâmetros e $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^T$ é o preditor linear;

3. Função de ligação: é uma função que relaciona a componente aleatória à componente sistemática, isto é,

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

sendo $g(\cdot)$ uma função monótona e diferenciável, que determina a escala em que a linearidade é suposta. Como os parâmetros do modelo $\beta_1, \dots, \beta_p$ não estão sujeitos à restrições, $g(\mu_i)$ podem assumir qualquer valor em $(-\infty, \infty)$, deste modo a forma da função apropriada é determinada em alguma escala pelo domínio de variação de $\mu_i = E(Y_i)$ (URBANO, 2012). Utiliza-se a seguinte expressão,

$$g(\mu_i) = \eta_i = \sum_j^p x_{ij}\beta_j,$$

para descrever a relação entre os componentes do modelo linear generalizado, em que $g(\cdot)$ é a função de ligação e η_i o preditor linear, sendo este, uma função linear dos parâmetros desconhecidos $\beta = (\beta_1, \dots, \beta_p)$ (CORDEIRO; PAULA, 1989).

2.3 Família exponencial

Sejam $Y_1, \dots, Y_n$, variáveis aleatórias independentes, cada uma com função de probabilidade para o caso discreto ou função densidade de probabilidade para o caso contínuo na forma dada abaixo,

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

sendo as funções $b(\cdot)$ e $c(\cdot)$ conhecidas, $\phi > 0$ interpretado como um parâmetro de dispersão, considerado conhecido, comum à distribuição de todas as observações contrariamente ao parâmetro natural θ_i que pode variar de observação para observação. Geralmente tem-se $a(\phi) = \phi$ (CORDEIRO; PAULA, 1989). Em (1), tem-se a família exponencial na forma canônica, com parâmetro canônico θ_i .

O valor médio e a variância de Y_i com distribuição pertencente à família em (1) são

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

sendo,

$$V_i = V(\mu_i) = \partial\mu_i/\partial\theta_i$$

denominada de função de variância, que é a parte da variância de Y_i que depende da média μ_i . A função de variância, descreve a possível dependência entre média e a variância (NELDER; LEE, 1991).

Como exemplos de distribuições pertencentes à família exponencial, tem-se a distribuição normal e a distribuição binomial.

Seja a variável aleatória $Y \sim N(\mu, \sigma^2)$. A função densidade de probabilidade de Y é expressa por

$$P(Y_i = y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(y - \mu)^2}{2\sigma^2} \right] \quad (2)$$

em que $-\infty < \mu < \infty$, $-\infty < y < \infty$ e $\sigma^2 > 0$.

Desenvolvendo-se (2), tem-se

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

e fazendo $\theta = \mu$, $\phi = \sigma^2$, $b(\theta) = \mu^2/2 = \theta^2/2$ e $c(y, \phi) = -\frac{1}{2}[\frac{y^2}{\phi} + \log(2\pi\sigma^2)]$, obtem-se (1).

Portanto, demonstra-se que a distribuição normal pertence à família exponencial de distribuições.

Se a variável aleatória $Y_i \sim \text{bin}(m_i, \pi_i)$, com probabilidade de sucesso π_i . A função de probabilidade de Y é expressa por

$$P(Y_i = y_i) = \binom{m_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{m_i - y_i}, \quad y_i = 0, 1, 2, \dots, m_i, \quad (3)$$

em que $0 < \pi_i < 1$ e m_i é um inteiro positivo

Desenvolvendo-se (3), tem-se

$$\begin{aligned} f(y_i; \pi_i) &= \exp \left[\log \binom{m_i}{y_i} + y_i \log(\pi_i) + (m_i - y_i) \log(1 - \pi_i) \right] \\ &= \left[y_i \log\left(\frac{\pi_i}{1 - \pi_i}\right) + m_i \log(1 - \pi_i) + \log \binom{m_i}{y_i} \right], \end{aligned} \quad (4)$$

e fazendo em (4) $\theta_i = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \log\left(\frac{\mu_i}{m_i - \mu_i}\right) \Rightarrow \mu = \frac{m \exp(\theta_i)}{1 + \exp(\theta_i)}$, $\phi = 1$, $b(\theta) =$

$-m_i \log(1 - \pi_i) = m_i \log(1 - \exp(\theta_i))$ e $c(y_i, \theta_i) = \log \binom{m_i}{y_i}$, obtem-se (1).

Portanto, a distribuição binomial pertence à família exponencial

2.4 Método de Estimação

Uma vez que o modelo linear generalizado tenha sido formulado, o passo seguinte é estimar os parâmetros e avaliar as estimativas. Para a estimação dos parâmetros

$(\beta_1, \dots, \beta_p)$ do modelo linear generalizado, um dos métodos utilizados é o da máxima verossimilhança. Este consiste em encontrar o vetor $\hat{\beta}(\hat{\beta}_1, \dots, \hat{\beta}_p)$ que maximiza a função de verossimilhança ou o que é usual o logaritmo da função de verossimilhança. O logaritmo da função de verossimilhança para n observações independentes como função de β para distribuições pertencentes à família em (1) é dado por (AGRESTI, 2002)

$$l(\beta) = \sum_{i=1}^n l_i = \sum_{i=1}^n a(\phi)^{-1} [y_i \theta_i - b(\theta_i)] + \sum_{i=1}^n c(y_i, \phi) \quad (5)$$

em que $\theta = q(\mu)$, $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = \sum_{j=1}^p x_{ij} \beta_j$, com ϕ conhecido.

As derivadas parciais de primeira ordem do logaritmo da função de verossimilhança são denominadas de vetor escore, dadas por

$$U_j = \frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^n \frac{(y_i - \mu_i)}{Var(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \quad (6)$$

em que

$$\frac{\partial l_i}{\partial \theta_i} = a(\phi)^{-1} [y_i - b'(\theta_i)], \quad \mu_i = b'(\theta_i) \Rightarrow \frac{\partial l_i}{\partial \theta_i} = a(\phi)^{-1} (y_i - \mu_i)$$

$$\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i) = \frac{Var(Y_i)}{a(\phi)} \Rightarrow \frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{a(\phi)^{-1} Var(Y_i)}$$

$$\eta_i = \sum_j x_{ij} \beta_j \Rightarrow \frac{\partial \eta_i}{\partial \beta_j} = x_{ij}$$

$$\eta_i = g(\mu_i), \quad \partial \mu_i / \partial \eta_i \text{ depende da função } g(\cdot)$$

As estimativas de máxima verossimilhança do vetor de parâmetros β são obtidas fazendo-se $\partial l(\beta) / \partial \beta_j = 0$, $j = 1, 2, \dots, p$. Em geral, as equações $\partial l(\beta) / \partial \beta_j = 0$ são não lineares e portanto não podem ser resolvidas explicitamente (CORDEIRO; PAULA, 1989). Nestes casos as equações são resolvidas numericamente mediante processo iterativo, como o método de Newton-Raphson ou o método escore (CORDEIRO; DEMÉTRIO, 2007)

2.4.1 Método de Newton-Raphson

Para solução do sistema de equações $\mathbf{U}_j = \mathbf{U}_\beta = \partial l(\beta) / \partial \beta_j = 0$, utiliza-se a versão multivariada do método de Newton-Raphson, cuja a expressão é da forma

$$\beta^{(m+1)} = \beta^{(m)} + (\mathbf{J}^{(m)})^{-1} \mathbf{U}^{(m)} \quad (7)$$

em que $\beta^{(m)}$ e $\beta^{(m+1)}$ são os vetores de parâmetros estimados nos passos m e $m+1$, $\mathbf{U}^{(m)}$ o vetor escore avaliado no passo m , com elementos $\partial l(\beta) / \partial \beta_j$ e $(\mathbf{J}^{(m)})^{-1}$ a inversa da

matriz de derivadas parciais de segunda ordem de $l(\boldsymbol{\beta})$, com elementos $-\partial^2 l(\boldsymbol{\beta})/\partial\beta_r\partial\beta_s$, avaliada no passo m (CORDEIRO; DEMÉTRIO, 2011)

2.4.2 Método escore de Fisher

O método escore de Fisher, usa a matriz de informação esperada de Fisher $\mathbf{K}$, ao invés da matriz de informação observada $\mathbf{J}$, ou seja, substitui a matriz de derivadas parciais de segunda ordem em (7) pela matriz de valores esperados das derivadas parciais. Esta conveniência deve ao fato da matriz $\mathbf{J}$ não ser positiva definida em algumas situações (PAULA, 2013). Logo,

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + (\mathbf{K}^{(m)})^{-1}\mathbf{U}^{(m)}, \quad (8)$$

em que $\mathbf{K}$ tem elemento típico dado por

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

Pré-multiplicando ambos os membros de (8) por $\mathbf{K}^{(m)}$, tem-se

$$\mathbf{K}^{(m)}\boldsymbol{\beta}^{(m+1)} = \mathbf{K}^{(m)}\boldsymbol{\beta}^{(m)} + \mathbf{U}^{(m)} \quad (9)$$

Para a obtenção de $k_{r,s}$ de $\mathbf{K}$ usa-se (6) e o seguinte resultado

$$E \left[\frac{\partial^2 l_i}{\partial\beta_r\partial\beta_s} \right] = -E \left[\frac{\partial l_i}{\partial\beta_r} \right] \left[\frac{\partial l_i}{\partial\beta_s} \right]. \quad (10)$$

que garante para as famílias exponenciais (COSX; HINKLEY, 1974 apud AGRESTI, 2002),

$$E \left[\frac{\partial^2 l_i}{\partial\beta_r\partial\beta_s} \right] = -E \left[\frac{(Y_i - \mu_i)x_{ir}}{Var(Y_i)} \frac{\partial\mu_i}{\partial\eta_i} \frac{(Y_i - \mu_i)x_{is}}{Var(Y_i)} \frac{\partial\mu_i}{\partial\eta_i} \right] = \frac{-x_{ir}x_{is}}{Var(Y_i)} \left(\frac{\partial\mu_i}{\partial\eta_i} \right)^2.$$

uma vez que $l(\boldsymbol{\beta}) = \sum l_i$, tem-se

$$k_{r,s} = E \left[-\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\beta_r\partial\beta_s} \right] = \sum_i^n \frac{x_{ir}x_{is}}{Var(Y_i)} \left(\frac{\partial\mu_i}{\partial\eta_i} \right)^2.$$

Generalizando a partir deste elemento típico para toda a matriz, a matriz de informação de Fisher tem a forma

$$\mathbf{K} = \mathbf{X}'\mathbf{W}\mathbf{X}, \quad (11)$$

em que $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$ é uma matriz diagonal de pesos com elementos na diagonal principal

$$w_i = \left(\frac{\partial\mu_i}{\partial\eta_i} \right)^2 / Var(Y_i).$$

A obtenção do vetor $\mathbf{U} = \mathbf{U}(\boldsymbol{\beta})$ decorre inicialmente multiplicando e dividindo a expressão em (6) por $\partial\mu_i/\partial\eta_i$, obtendo

$$\mathbf{U} = \mathbf{U}_{\boldsymbol{\beta}} = \phi \sum_{i=1}^n \frac{(y_i - \mu_i)}{V(\mu_i)} \left(\frac{\partial\mu_i}{\partial\eta_i} \right)^2 \frac{\partial\eta_i}{\partial\mu_i} x_{ij} \quad (12)$$

sendo sua forma matricial expressa por

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

em que $\mathbf{G} = \text{diag} \left\{ \frac{\partial\eta_1}{\partial\mu_1}, \dots, \frac{\partial\eta_n}{\partial\mu_n} \right\} = \text{diag}[g'(\mu_1), \dots, g'(\mu_n)]$, sendo $g(\cdot)$ a função de ligação. Substituindo $\mathbf{U}$ e K pelas suas respectivas formas matriciais em (9) e eliminando ϕ , tem-se

$$\begin{aligned} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} &= \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m)} + \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{G}^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)}) \\ &= \mathbf{X}^T \mathbf{W}^{(m)} \left[\mathbf{X} \boldsymbol{\beta}^{(m)} + \mathbf{G}^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)}) \right] \\ \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} &= \mathbf{X}^T \mathbf{W}^{(m)} \left[\boldsymbol{\eta}^{(m)} + \mathbf{G}^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)}) \right] \end{aligned} \quad (13)$$

Define-se, a variável dependente ajustada da forma $\mathbf{z}^{(m)} = \boldsymbol{\eta}^{(m)} + \mathbf{G}^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)})$ e substituímos em (13), tem-se

$$\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} = \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)} \quad (14)$$

ou ainda, pré-multiplicando (14) por $(\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1}$ em ambos os lados, obtem-se

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)} \quad (15)$$

Segundo Cordeiro e Demétrio (2011) o processo iterativo em (15) é válido para qualquer modelo linear generalizado, não depende do parâmetro de dispersão, as funções de variância e de ligação entram no processo por meio de $\mathbf{W}$ e de $\mathbf{z}$ e os z_i ajustados são não correlacionados.

Segundo Agresti (2002), a função de verossimilhança para modelos lineares generalizados também determina a matriz de variância e covariância assintótica do estimador $\hat{\boldsymbol{\beta}}$ de máxima verossimilhança. Esta matriz é o inverso da matriz de informação $\mathbf{K}$. Assim, a matriz de covariância assintótica de $\hat{\boldsymbol{\beta}}$ é estimada por

$$\widehat{cov}(\hat{\boldsymbol{\beta}}) = \hat{\mathbf{K}}^{-1} = (\mathbf{X}' \widehat{\mathbf{W}} \mathbf{X})^{-1} \quad (16)$$

em que $\widehat{\mathbf{W}}$ e $\mathbf{W}$ avaliado em $\hat{\boldsymbol{\beta}}$.

2.4.3 Critério de convergência

Segundo Cordeiro e Demétrio (2011), o critério do desvio é o mais usado para verificar convergência, este consiste em verificar se $|desvio^{m+1} - desvio^m| < \xi$, sendo o desvio definido em (20). Ainda, apresentam um outro critério para verificar a convergência do algoritmo iterativo em (15), dado por,

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

sendo ξ um número positivo e suficientemente pequeno.

2.5 Medidas da qualidade do ajuste

Segundo Cordeiro e Paula (1989), a etapa de inferência tem por objetivo principal verificar a precisão do modelo como um todo e realizar um estudo detalhado quanto às discrepâncias. Estas discrepâncias, quando significativas, podem implicar na escolha de um outro modelo.

Rao e Wu (2005) argumentam que a escolha deve recair sobre o modelo que mais se aproxima do modelo verdadeiro a partir de um conjunto de modelos candidatos, sendo sua adequação avaliada pela sua capacidade de apresentar pequenas discrepâncias entre os valores reais e seus respectivos valores. O modelo também deve ser parcimonioso, ou seja, o número de parâmetros deve ser o menor possível.

Segundo McCullagh e Nelder (1989), o ajustamento de um modelo a um conjunto de dados observados y_i pode ser encarado como uma maneira de substituir y_i por um conjunto de valores estimados $\hat{\mu}_i$ para um modelo com poucos parâmetros. Esses valores não serão exatos, logo é necessário definir um limite para essa discrepância.

Para Cordeiro e Demétrio (2011), as n observações podem ser ajustados modelos contendo até n parâmetros. O modelo que tem um único parâmetro é denominado de modelo nulo, representado por um vetor μ comum a todas as observações. O modelo contendo n parâmetros especificado pelas médias $\mu_1, \dots, \mu_n$ linearmente independente é denominado de modelo saturado ou completo, sendo que o mesmo atribui toda a variação dos dados à componente sistemática, ajustando-se desta forma perfeitamente, reproduzindo assim os próprios dados. Conceitua ainda, modelo minimal como sendo aquele que tem o menor número de termos possíveis para o ajuste e o modelo maximal, aquele que considera a inclusão do maior número de termos no ajuste e qualquer modelo com p

parâmetros linearmente independentes, localizado entre o modelo minimal e o maximal é denominado de modelo sob pesquisa ou corrente.

Nelder e Wedderburn (1972) introduziram no contexto dos modelos lineares generalizados, uma medida de discrepância, denominada de deviance, traduzida por Cordeiro (1986) apud Cordeiro e Demétrio (2011) como desvio, que é uma medida da qualidade do ajuste de um MLG. Um modelo mal ajustado apresenta um grande desvio, enquanto um modelo bem adequado apresenta um pequeno desvio. Cordeiro e Demétrio (2011) apresentam a seguinte expressão para a função desvio

$$S_p = 2(\hat{l}_n - \hat{l}_p) \quad (17)$$

Sendo $\hat{l}_n$ e $\hat{l}_p$ os máximos do logaritmo da função de verossimilhança para os modelos saturado e corrente, dados por

$$\hat{l}_n = \phi^{-1} \sum_{i=1}^n [(y_i)\tilde{\theta}_i - b(\tilde{\theta}_i)] + \sum_{i=1}^n c(y_i, \phi) \quad (18)$$

e

$$\hat{l}_p = \phi^{-1} \sum_{i=1}^n [(y_i)\hat{\theta}_i - b(\hat{\theta}_i)] + \sum_{i=1}^n c(y_i, \phi) \quad (19)$$

sendo $\tilde{\theta} = q(y_i)$ e $\hat{\theta}_i$ as estimativas de máxima verossimilhança dos parâmetros canônicos sob os modelos saturado e corrente. Substituindo (18) e (19) em (17), tem-se

$$S_p = \phi^{-1} D_p = 2\phi^{-1} \sum_{i=1}^n [y_i(\tilde{\theta}_i - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i)] \quad (20)$$

em que $S_p = D_p/\phi$ é denominado de desvio escalonado e D_p de desvio.

O desvio ainda pode ser decomposto como

$$S_p = \phi^{-1} \sum_{i=1}^n d_i^2, \quad (21)$$

sendo que d_i , mede a diferença dos logaritmos das funções de verossimilhanças observada e ajustada para cada observação i correspondente. Este é denominado de componente do desvio. A soma d_i^2 mede a discrepância total entre os dois modelos. Paula (2013) argumenta que um valor pequeno para a função desvio indica que, para um número menor de parâmetros, obtem-se um ajuste tão bom quanto o ajuste com o modelo saturado.

Para o teste, compara-se o desvio e seus graus de liberdade $v = n - p$ com alguma distribuição teórica de probabilidade, sendo n o número de observações e

p o posto da matriz do modelo sob pesquisa. Geralmente, adota-se a distribuição qui-quadrado. A dificuldade em realizar este teste é que para alguns modelos, o parâmetro ϕ é desconhecido. Quando o modelo é verdadeiro, o desvio não é, em geral distribuído como uma qui-quadrado (χ^2_{n-p}), nem mesmo assintoticamente. Apesar disso, contenta-se em testar um MLG sem muito rigor, comparando o desvio com os percentis da distribuição $\chi^2_{n-p, \alpha}$ (CORDEIRO; PAULA, 1989).

Assim, quando $S_p = \phi^{-1}D_p \leq \chi^2_{n-p, \alpha}$, pode-se considerar que existe evidências, a um nível aproximado de $100\alpha\%$ de significância que o modelo proposto está bem ajustado aos dados. Ou ainda, se o valor de D_p for próximo de $n - p$ (graus de liberdade) de uma distribuição χ^2_{n-p} pode ser um indicativo de um bom ajuste (CORDEIRO; DEMÉTRIO, 2011).

Segundo Cordeiro e Demétrio (2011), uma outra medida da discrepância do ajuste de um modelo a um conjunto de dados é a estatística de Person generalizada X_p^2 , cuja a expressão é

$$X_p^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}, \quad (22)$$

Sendo $V(\hat{\mu}_i)$ a função de variância estimada sob o modelo que está sendo ajustado. A deviance leva vantagem como medida de discrepância, visto que ela é aditiva para conjuntos encaixados de modelos, entretanto $\mathbf{X}_p^2$ é preferida em algumas situações devido a sua interpretação mais direta (McCULLAG; NELDER, 1989). No caso da distribuição binomial e Poisson, sendo $\phi = 1$, X_p^2 é a conhecida estatística de Pearson, cuja a forma

$$X_p^n = \sum_{i=1}^n \frac{(O_i - e_i)^2}{e_i}$$

sendo e_i a frequência esperada e O_i a frequência observada. Cabe salientar, que toda inferência feita para os MLGs é baseada em resultados assintóticos. Conforme Cordeiro e Paula (1989), quando a amostra é pequena, pouco se sabe sobre a validade desses resultados. Assim, tanto o desvio quanto a estatística de Pearson generalizada possuem distribuição assintoticamente normal.

2.5.1 Análise de desvio

A análise de desvio “Analysis of the Deviance” – ANODEV é uma generalização da análise de variância para os MLGs e que permite avaliar a contribuição de

cada termo no modelo final. Segundo Cordeiro e Paula (1989), a análise de desvio tem como objetivo a construção de uma sequência de modelos encaixados e a verificação da significância dos termos adicionais. Entretanto, deve-se ter atenção com a (ANODEV) devido a não ortogonalidade dos termos que em geral ocorre. Assim, é necessário considerar diferentes sequências para o modelo, pois cada uma produzirá uma tabela ANODEV diferente.

Coedeiro e Demétrio (2011) apresentam o processo da construção da ANODEV, para uma sequência de modelos encaixados, $M_{p1}, \dots, M_{pr}$ de dimensões $p_1 < p_2 < \dots < p_r$ obtidos pela adição de termos um a um, com matrizes dos modelos $X_{n_{p1}}, X_{n_{p2}}, \dots, X_{n_{pr}}$, com correspondente sequência de desvio decrescente $D_{p1} > D_{p2} > \dots > D_{pr}$, tendo os modelos a mesma distribuição e a mesma função de ligação. A Tabela 1 apresenta um exemplo de ANODEV para um experimento inteiramente casualizado, com r repetições e tratamento no esquema fatorial, com a níveis do fator A e b níveis do fator B.

Tabela 1 – Construção da análise de desvio - ANODEV, para modelos lineares generalizados

Modelo	g.l	Desvio	Dif. de desvio	Dif.de g.l
Nulo	$rab - 1$	D_1		
A	$a(rb - 1)$	D_A	$D_1 - D_A$	$a - 1$
A+B	$a(rb - 1) - (b - 1)$	D_{A+B}	$D_A - D_{A+B}$	$b-1$
A+B+AB	$ab(r - 1)$	D_{A*B}	$D_{A+B} - D_{A*B}$	$(a - 1)(b - 1)$
Saturado	0	0	D_{AB}	$ab(r - 1)$

Fonte: Cordeiro e Demétrio - 2011.

Segundo Cordeiro e Paula (1989) para os modelos encaixados M_q e M_p ($M_q \subset M_p, q < p$) a estatística $(D_q - D_p)$, com $p - q$, graus de liberdade é interpretada como uma medida da variação dos dados explicada pelos termos que estão no modelo M_p , e não estão no modelo M_q , incluídos os efeitos dos termos em M_q e ignorando quaisquer efeitos dos termos que não estão em M_p . Assintoticamente para ϕ conhecido, se

$$(S_q - S_p) = \phi^{-1}(D_q - D_p) > \chi_{p-q}^2, \alpha \quad (23)$$

os efeitos dos termos que estão em M_p e não estão em M_q são significativos. Assim, para

uma sequência de M_{pk} modelos encaixados, pode-se calcular a deviance e proceder aos testes de significância, construindo a tabela ANODEV (VIEIRA, 2004).

Segundo Cordeiro e Demétrio (2011) se ϕ é desconhecido, deve-se obter uma estimativa $\hat{\phi}$ consistente, de preferência baseada no modelo maximal (com m parâmetros), e a inferência pode ser baseada na estatística F , expressa por

$$F = \frac{(D_q - D_p)/(p - q)}{\hat{\phi}} \sim F_{(p-q, n-m)} \quad (24)$$

sendo D_p o desvio do modelo m_p com p parâmetros e D_q o desvio do modelo m_q com q parâmetros e $\hat{\phi}$ uma estimativa razoável de ϕ .

Para estimação de ϕ , pode-se utilizar o método do desvio que para um modelo bem ajustado as observações, espera-se que o desvio escalonado S_p tenha valor esperado igual a $n - p$. Assim a estimativa do parâmetro ϕ é dada por

$$\hat{\phi}_d = \frac{D_p}{n - p} \quad (25)$$

Paula (2013), apresenta um estimador consistente para ϕ (de momentos) que não envolve o processo iterativo, é baseado na estatística de Pearson generalizada ($\mathbf{X}_p^2$), sendo sua expressão da forma,

$$\hat{\phi}_p = \frac{1}{n - p} \sum_{i=1}^n \left\{ \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \right\}$$

2.6 Análise de resíduos e diagnósticos

Segundo Cordeiro e Paula (1989), por técnica de diagnóstico, entende-se a análise dos resíduos para detectar observações aberrantes e o estudo da influência de observações sobre o ajustamento global do modelo. Usualmente, nos modelos clássicos de regressão, os resíduos são utilizados como técnicas para verificar : *i*) violações nas suposições de homogeneidade das variâncias ou de normalidade dos resíduos; *ii*) presença de valores atípicos e *iii*) influência de observações individuais no ajustamento global do modelo. Esse tipo de análise só é adequado quando a variância das observações é constante, suposição esta que não é necessária em aplicações de MLGs (McCULLAGH; NELDER, 1989).

Os resíduos, no âmbito dos MLGs, são utilizados para verificar a adequação do modelo ajustado em relação à escolha da função de variância, da função de ligação e dos termos do preditor linear. Além disso, os resíduos são também úteis para iden-

tificar a presença de pontos atípicos, que podem ser influentes ou não no modelo final (CORDEIRO; LIMA NETO, 2004).

Segundo Cordeiro e Demétrio (2011), os tipos de resíduos mais comuns nos modelos lineares generalizados são:

a) Resíduo de Pearson

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{V}_i}} \quad (26)$$

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

b) Resíduo de Pearson estudentizado

$$r_i^{P'} = \frac{(y_i - \hat{\mu}_i)}{\sqrt{V(\hat{\mu}_i)(1 - \hat{h}_{ii})}} \quad (27)$$

Sendo $\hat{h}_{ii}$, o i -ésimo elemento da diagonal principal da matriz de projeção $\mathbf{H}$, expressa por,

$$\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}, \quad (28)$$

é, em geral, assimétrico, mesmo para grandes amostras.

c) Resíduo de Anscombe

$$A_i = \frac{N(y_i) - N(\hat{\mu}_i)}{N'(\hat{\mu}_i) \sqrt{\hat{V}(\hat{\mu}_i)}} \quad (29)$$

Segundo Paula (2013), a distribuição dos resíduos de Anscombe pode estar mais próximo da normalidade. Sendo $N(\cdot)$ uma transformação utilizada para normalizar a distribuição de Y . Para os MLGs, essa transformação é da forma,

$$N(\mu) = \int_0^\mu V^{-1/3}(\mu) d\mu \quad (30)$$

d) Resíduo componente do desvio

Definidos como iguais as raízes quadradas dos componentes do desvio com sinal igual ao sinal da diferença $(y_i - \hat{\mu}_i)$, ou seja

$$r_i^D = \text{sin}al(y_i - \hat{\mu}_i) \sqrt{2}(\hat{l}_n - \hat{l}_p)^{1/2} \quad (31)$$

sendo $\hat{l}_n$ e $\hat{l}_p$ os máximos do logaritmo da função verossimilhança para os modelos saturados e corrente, respectivamente, para cada observação i .

Um valor grande para r_i^D indica que a i -ésima observação é mal ajustada pelo modelo. Conforme Cordeiro e Demétrio (2011), as vantagens desse resíduo é que não requer o conhecimento da função normalizadora e o fato de ser definido para toda a observação, desde que estas forneçam uma contribuição para o logaritmo da função de verossimilhança. Observa-se que, para modelos bem ajustados, as diferenças entre r_i^D e $r_i^{P'}$ são pequenas enquanto que os resultados com uso de r_i^D e A_i são similares.

f) Resíduo componente do desvio estudentizado

$$r_i^{D'} = \frac{r_i^D}{\sqrt{(1 - \hat{h}_{ii})}} \quad (32)$$

Williams (1984) apud Paula (2013) mostra por meio de simulações que a distribuição de $r_i^{D'}$ tende a estar mais próxima da normalidade do que as demais distribuições dos demais resíduos.

McCullagh e Nelder (1989) recomendam plotar os resíduos componente do desvio contra os valores ajustados ou ainda, plotar os valores absolutos dos resíduos componente do desvio contra os valores ajustados, tendência no gráfico, é um indicativo da má escolha da função de variância. Ainda, recomendam a utilização do gráfico de probabilidade normal dos resíduos componente do desvio. Cordeiro e Paula (1989), afirma que um gráfico de resíduos padronizados *versus* os valores ajustados que não apresenta tendência pode ser um indicativo de que a relação funcional variância/média proposta para os dados é satisfatória.

Lee e Nelder (1998) afirmam que os resíduos dados pelos componentes do desvio podem ser considerados como aproximadamente normais, com média zero e variância constante. Assim, recomendam o uso de dois tipos de gráficos: (i) resíduos estudentizados *versus* valores ajustados e (ii) resíduos absolutos *versus* valores ajustados. Sendo que esses gráficos são interpretados como no caso da regressão clássica.

McCullagh e Nelder (1989) sugerem o uso do gráfico normal de probabilidades (*normal plot*) para os resíduos e o semi-normal de probabilidades (*half normal plot*) para medidas positivas como é o caso de h ou a distância de Cook modificada. No caso do gráfico normal de probabilidade para resíduos, espera-se que, na ausência de pontos discrepantes, o aspecto seja linear.

Paula (2013) apresenta vários algoritmos implementados no programa R para a construção de gráficos de probabilidades com envelopes de simulações, dentre eles,

o gráfico de probabilidade normal padrão com envelope de simulação *versos* os resíduos componentes do desvio para o MLG binomial.

2.7 Adequação da função de ligação

A adequação da função de ligação pode ser verificada informalmente por meio do gráfico dos resíduos estudentizados *versus* valores ajustados. Caso esse gráfico apresente algum padrão (comportamento não aleatório) e a linha resultante do amortecimento (*lowess*) não sendo aproximadamente horizontal e próxima à linha reta horizontal de ordenada zero, há indícios de que a função de ligação não seja adequada (VIEIRA, 2004).

Cordeiro e Demétrio (2011) apresentam dois métodos formais para verificar a adequacidade da função de ligação: i) adicionar $\Delta = \boldsymbol{\eta} \otimes \boldsymbol{\eta}$ como uma variável explicatória e examinar a mudança ocorrida no desvio. Se ocorrer uma diminuição considerada elevada há evidências de que a função de ligação é insatisfatória, pode se usar o teste da razão de verossimilhança ou o teste de escore e ii) indexar a família de ligação por um parâmetro λ e fazer um teste da hipótese $H_0 : \lambda = \lambda_0$, usando-se os teste de razão de verossimilhança e escore, e que a não rejeição de H_0 , indica que a função de ligação é adequada.

2.8 Adequação da função de variância

A adequação da função de variância pode ser verificada informalmente por meio do gráfico dos resíduos absolutos estudentizados *versus* os valores ajustados, sendo o padrão nulo para esse gráfico uma distribuição aleatória de média zero e amplitude constante. Cordeiro e Demétrio (2011) apresentam um método formal para verificar a adequação da função de variância, este consiste em indexar essa função por um parâmetro λ e fazer um teste de hipótese $H_0 : \lambda = \lambda_0$, usando-se os testes da razão de verossimilhança e escore .

3 MATERIAL E MÉTODOS

3.1 Material

Os ovos de *Aedes aegypti* usados durante os testes de viabilidade foram coletados de seis gaiolas da colônia do laboratório de malária e dengue do Instituto Nacional de Pesquisa do Amazonas - INPA, nas quais diariamente eram colocados copos plásticos de volume de 100 ml contendo 20 ml de água, com as laterais forradas com uma faixa de papel filtro medindo 3 cm de altura por 22 cm de comprimento, para servir de substrato de postura dos ovos. Cada copo permanecia dentro da gaiola por duas horas, sempre no período da tarde. Após serem recolhidas, as cartelas com os ovos eram expostas à umidade durante 72 horas para que fosse finalizado o desenvolvimento do embrião.

As cartelas eram deixadas por 24 horas, no ambiente até ficarem completamente secas. A seguir, fazia-se a contagem e o armazenamento nos respectivos recipientes, envelopes de papel (tipo ofício médio), sacos plástico (200 ml) e copos plástico de (100 ml) abertos (Figura 1). Os ovos permaneceram armazenados até completarem os respectivas períodos de armazenamento (idade), os quais foram definidos previamente pelos pesquisadores. Assim, os ovos ficaram armazenados por períodos de 12, 19, 32, 61, 89, 118, 158, 186 dias, momento em que foram tratados com água (estimulados) e a viabilidade foi medida por meio dos totais de larvas vivas provenientes dos totais de ovos usados em cada período e tipo de armazenamento.

3.1.1 Testes de viabilidade

Para estimular os ovos, mergulharam-se as cartelas em bacias plásticas, com 1,5 ml de água de poço artesiano, acrescido de 10 ml de pó de fígado, usado como alimento para as larvas (SCARPASSA; TADEI, 1990). Aleatoriamente, foram selecionadas duas cartelas entre cinco, de cada idade armazenada, sendo dezesseis bacias por recipiente, totalizando 48 bacias. A contagem da quantidade de larvas presentes nas cubas deu-se diariamente durante um período de 10 dias, verificando-se no sexto dia uma frequência de larvas quase nula e não mais ocorrência após o décimo dia. Esse tempo de observação foi determinado com base em um experimento piloto em que foram feitas leituras de até 25 dias. A reposição da água era feita de acordo com a evaporação e a limpeza das bacias realizadas duas vezes por semana.

Os ovos ficaram em meio à mata, em uma gaiola de madeira telada no interior de uma cabana, com proteção do sol, chuva e dos possíveis predadores (Figura 2).

Figura 1 – Tipos de recipientes em que os ovos de *Aedes aegypti* foram armazenados: envelopes de papel, copos plástico e sacos plástico.

Figura 2 – Local rodeada por vegetação e com cobertura em que os lotes dos ovos de *Aedes aegypti* foram mantidos por diferentes períodos de armazenamento

3.2 O experimento

O experimento consiste de um delineamento inteiramente casualizado, cuja a variável resposta Y_i^* é a proporção de sucesso (número de larvas provenientes de totais de m_i ovos/totais de m_i ovos). As variáveis explanatórias ou explicativas, representadas por recipientes, classificada como qualitativa com três níveis (copo plástico, envelope de papel e saco plástico) e a variável período de armazenamento, que pode ser considerada tanto do tipo qualitativa como do tipo quantitativa. Neste trabalho, considerou-se esta do tipo quantitativa, com duas repetições por recipiente.

Pode-se observar, na Tabela 2, parte dos resultados referentes à viabilidade dos ovos de *Aedes aegypti* produzidos em condições laboratoriais segundo tipo de recipientes e períodos de armazenamento. Os dados aqui apresentados, foram gentilmente cedidos para este trabalho pelo professor Dr. Wanderli Pedro Tadei do Instituto Nacional de Pesquisa do Amazonas - INPA.

Tabela 2 – Frequência de larvas e total dos ovos de *Aedes aegypti* produzidos em condições laboratoriais, segundo tipo de recipientes e período de armazenamento

Recipientes	Idade	Número de larvas	Total de ovos
Copo	12	127	127
Copo	19	102	102
⋮	⋮	⋮	⋮
Copo	186	4	149
Envelope	12	100	104
Envelope	19	72	110
⋮	⋮	⋮	⋮
Envelope	186	1	176
Saco	12	180	186
Saco	19	103	123
⋮	⋮	⋮	⋮
Saco	186	0	130

Fonte: Instituto Nacional de Pesquisa do Amazonas - INPA.

3.3 Métodos

Para a análise estatística do experimento, ajustaram-se modelos sob a metodologia dos modelos lineares generalizados (McCULLANGH; NELDER, 1989).

3.4 Modelo linear generalizado binomial

Admitindo que Y_i^* é a proporção de sucesso, cada um com probabilidade de ocorrência (π_i), temos que o modelo natural para esta situação é o MLG binomial (McCULLANGH; NELDER, 1989; DOBSON, 2002; CORDEIRO; DEMÉTRIO, 2011). Como funções de ligação, optaram-se pelas ligações logítisca e probito, que são transformações do valor esperado. Como parte sistemática, tem-se um delineamento inteiramente casualizado com fatores recipiente (qualitativo) com três níveis e período de armazenamento do ovo (quantitativo). Os preditores lineares considerados são retas que podem ser concorrentes ou paralelas. Assim, o MLG binomial é da forma:

$$\begin{aligned}
 m_i Y_i^* &\sim \text{bin}(m_i, \pi_i), \quad i = 1, 2, \dots, n \\
 \mu_i &= E(Y_i^*) = \pi_i \\
 \boldsymbol{\eta} &= \mathbf{X}_j \boldsymbol{\beta}_j \\
 \eta_i &= \log \left(\frac{\pi_i}{1 - \pi_i} \right) = \log \left(\frac{\mu_i}{m_i - \mu_i} \right)
 \end{aligned} \tag{33}$$

$$\eta_i = \Phi^{-1}(\mu_i) \tag{34}$$

em que $m_i Y_i^*$ é o número de sucessos;

$\mathbf{Y}^*$ é o vetor de proporções observadas, de dimensão $n \times 1$;

$\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$ é o vetor dos preditores lineares;

(33) - função de ligação logística;

(34) - função de ligação probito;

$\mathbf{X}$ é a matriz do delineamento, de dimensões $n \times p$;

$\boldsymbol{\beta}$ é o vetor de parâmetros p desconhecidos, de dimensões $p \times 1$;

m o número de ensaios independentes.

Os preditores lineares retas concorrentes e retas paralelas, tem expressão da forma respectivamente,

$$\eta_{ij} = \alpha_j + \beta_j x_i \tag{35}$$

$$\eta_{ij} = \alpha_j + \beta x_i \quad (36)$$

com $i = 1, 2, \dots, 8$, períodos de armazenamentos e $j = 1, 2, 3$, tipos de recipientes.

3.5 Modelos ajustados

Inicialmente foram ajustados quatro MLGs para o conjunto de dados apresentados na Tabela 2, sendo consideradas as funções de ligação logística e probito, com preditores lineares retas concorrentes e retas paralelas. Entretanto, como foi constatada falta de ajuste dos modelos iniciais ajustados, procederam-se novos ajustes, com inclusão de um peso ($w = 1/\hat{\phi}_p$) na variável binomial. Detalhes sobre o procedimento de inclusão de peso podem ser encontrados em (FOX, 2002) e (MENÉNDEZ FERNÁNDEZ, 2013). Também, foram realizados ajustes em que optou-se por retirar duas observações do conjunto de dados. Paula (2013) e Fox (2002) discutem resultados em que uma ou duas observações são eliminadas do conjunto de dados com o propósito de melhorar a qualidade do ajuste.

3.6 Modelo binomial

Para determinar as equações de verossimilhança e posteriormente estimá-las é necessário deduzir o modelo binomial para dados na forma de proporção. Seja Y^* a proporção de sucesos em m ensaios independentes, cada um com probabilidade de ocorrência π . Assumindo que $mY \sim \text{bin}(m, \pi)$. A função de probabilidade de Y^* para a observação i fica expressa na forma

$$\begin{aligned} f(y_i^*; \pi_i, m_i) &= \binom{m_i}{m_i y_i^*} \pi_i^{m_i y_i^*} (1 - \pi_i)^{m_i - m_i y_i^*} \\ &= \binom{m_i}{m_i y_i^*} \left(\frac{\pi_i}{1 - \pi_i} \right)^{m_i y_i^*} (1 - \pi_i)^{m_i} \\ &= \exp \left\{ \frac{y_i^* \log\left(\frac{\pi_i}{1 - \pi_i}\right) + \log(1 - \pi_i)}{\frac{1}{m_i}} + \log \left(\binom{m_i}{m_i y_i^*} \right) \right\} \end{aligned} \quad (37)$$

Fazendo $\theta_i = \log \left(\frac{\pi_i}{1 - \pi_i} \right)$, $b(\theta_i) = -\log[1 + \exp(\theta_i)]$, $a(\phi) = \frac{1}{m_i}$ e $c(y_i, \phi) = \log \left(\binom{m_i}{m_i y_i} \right)$, obtem-se,

$$f(y_i, \theta, \phi) = \exp \left\{ \frac{y_i \theta - \log(1 + \exp(\theta))}{\frac{1}{m_i}} + \log \left(\binom{m_i}{m_i y_i} \right) \right\}, \quad (38)$$

o que mostra que a distribuição binomial para dados na forma de proporção é membro da família exponencial de distribuições.

Assim, a média de (38) é dada por,

$$\mu_i = E(Y_i^*) = b'(\theta_i) = \frac{\exp(\theta_i)}{(1 + \exp(\theta_i))} \Rightarrow \mu_i = \frac{\frac{\pi_i}{1 - \pi_i}}{1 + \frac{\pi_i}{1 - \pi_i}} = \pi_i$$

a variância,

$$Var(Y_i^*) = a(\phi)b''(\theta_i) = \frac{\exp(\theta_i)}{[1 + \exp(\theta_i)]^2 m_i} = \pi_i(1 - \pi_i)/m_i \quad (39)$$

é a função de variância,

$$V(\mu_i) = \pi_i(1 - \pi_i).$$

3.7 Equações de verossimilhança do MLG binomial

Lembrando a estimativa do vetor de parâmetros β pelo método da máxima verossimilhança, é determinado a partir do logaritmo da função de verossimilhança para n observações independentes, conforme descrito em (5). Assim, para o modelo binomial para dados em forma de proporção, o logaritmo da função de verossimilhança de (38) é:

$$l(\theta_i, \phi; y_i^*) = \sum_{i=1}^n m_i [y_i^* \theta_i - \log(1 + \exp(\theta_i))] + \sum_{i=1}^n \log \left(\frac{m_i}{m_i y_i^*} \right) \quad (40)$$

e o correspondente logaritmo da função de verossimilhança como função de β , é expresso por,

$$l(\beta) = \sum_{i=1}^n l_i = \sum_{i=1}^n m_i [y_i^* \theta_i - \log(1 + \exp(\theta_i))] + \sum_{i=1}^n \log \left(\frac{m_i}{m_i y_i^*} \right) \quad (41)$$

Como o vetor escore é formado pelas derivadas parciais de primeira ordem do logaritmo da função de verossimilhança, pode-se calcular pela regra da cadeia, o vetor escore $U(\beta) = \partial l(\beta) / \partial \beta$ de dimensão p , com elemento típico

$$U_j = \frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l_i}{\partial (\beta_j)} = \sum_{i=1}^n \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}, \forall_j \quad (42)$$

Aplicando a regra da cadeia para uma observação em (42), obtém-se

$$\frac{\partial l_i}{\partial \theta_i} = m_i \left[y_i^* - \frac{\exp(\theta_i)}{(1 + \exp(\theta_i))} \right], \quad \mu_i = b'(\theta_i) \Rightarrow \frac{\partial l_i}{\partial \theta_i} = m_i [y_i^* - \mu_i] \quad (43)$$

$$\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i) = \frac{Var(Y_i^*)}{a(\phi)} \Rightarrow \frac{\partial (\theta_i)}{\partial \mu_i} = \frac{1}{a(\phi)^{-1} Var(Y_i^*)} = \frac{1}{m_i Var(Y_i^*)} \quad (44)$$

$$\eta_i = \sum_j x_{ij} \beta_j \Rightarrow \frac{\partial \eta_i}{\partial \beta_j} = x_{ij} \quad (45)$$

$$\eta_i = g(\mu_i) \text{ e } \frac{\partial \mu_i}{\partial \eta_i}, \text{ depende da função } g(.) \quad (46)$$

usando os resultados (43), (44), (45) e (46), substituindo em (42) e igualando a zero, tem-se

$$\sum_{i=1}^n \frac{m_i(y_i^* - \mu_i)}{m_i \text{Var}(Y_i^*)} x_{ij} \frac{\partial \mu_i}{\partial \eta_i} = \frac{(y_i^* - \mu_i)}{\text{Var}(Y_i^*)} x_{ij} \frac{\partial \mu_i}{\partial \eta_i} = 0, \quad j = 1, \dots, p \quad (47)$$

Para o caso em que a função de ligação usada é a logística, tem-se:

$$\eta_i = \log \left(\frac{\pi_i}{1 - \pi_i} \right) = \alpha + \beta x_i, \quad (48)$$

e com algumas substituições algébricas em (48), chega-se a uma expressão de μ_i em função de η_i , o que possibilita calcular $\partial \mu_i / \partial \eta_i$. Assim,

$$\exp(\eta_i) = \left(\frac{\mu_i}{1 - \mu_i} \right) \Rightarrow \exp(\eta) - \exp(\eta_i) \mu_i = \mu_i \Rightarrow \mu_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad (49)$$

em seguida, derivando o resultado em (49) em relação a η_i

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{\exp(\eta_i)[1 + \exp(\eta_i)] - \exp(\eta_i) \exp(\eta_i)}{[1 + \exp(\eta_i)]^2} \Rightarrow \frac{\exp(\eta_i)}{[1 + \exp(\eta_i)]^2} \quad (50)$$

e substituindo (39) e o resultado de (50) em 47, tem-se

$$\sum_{i=1}^n \frac{(y_i - \mu_i)}{\frac{\exp(\eta_i)}{m_i[1 + \exp(\eta_i)]} \left[1 - \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right]} x_{ij} \frac{\exp(\eta_i)}{[1 + \exp(\eta_i)]^2} = 0$$

$$\sum_{i=1}^n m_i \frac{(y_i - \mu_i)[1 + \exp(\eta_i)]^2}{\exp(\eta_i)} x_{ij} \frac{\exp(\eta_i)}{[1 + \exp(\eta_i)]^2} = \sum_{i=1}^n m_i (y_i - \mu_i) x_{ij} = 0, \quad (51)$$

assim, o resultado em (51) são as equações de verossimilhança do modelo linear generalizado binomial com função de ligação logística (AGRESTI, 2002).

Para o caso em que a função de ligação usada é a probito, tem-se a forma

$$\sum_i \frac{m_i(y_i - \pi_i)}{\pi_i(1 - \pi_i)} \phi \left(\sum_j \beta_j x_{ij} \right) = 0 \quad (52)$$

As deduções necessárias e argumentações para se determinar a expressão (52), podem ser encontradas em (AGRESTI, 2002).

3.8 Ajuste dos modelos

O processo para o ajuste dos modelos deu-se mediante solução das equações de verossimilhanças, por meio da função *glm* do programa estatístico R (R DEVELOPMENT CORE TEAM, 2013). Para o ajuste no R por meio da função *glm*, deve-se especificar a *formula* (a definição do modelo) e *family* (a distribuição assumida pela variável resposta com a função de ligação a ser utilizada). Por exemplo,

$$modelo <- glm(y \sim x_1 + x_2, family = binomial)$$

Se a função de ligação usada for diferente do default, basta especificar a função de ligação desejada através do comando *link*. Por exemplo,

$$modelo <- glm(y \sim x_1 + x_2, family = binomial(link = "probit"))$$

O comando *summary(modelo)* dá um resumo do resultado do ajuste. Pode ser encontrado em Crawley (2007) uma explanação mais detalhada sobre a função *glm* e seus argumentos utilizando o programa R.

3.9 Desvio do MLG binomial

Para avaliar a qualidade do ajuste dos modelos, utilizou-se o desvio residual (*deviance*), que para um modelo bem ajustado aos dados, deve-se esperar que seja próximo ao número de graus de liberdade do modelo (CORDEIRO ; PAULA, 1989). Para o teste de ajuste do modelo corrente, compara-se o desvio com o valor crítico da distribuição qui-quadrado à um nível de significância igual a α . Se este for maior que $\chi^2_{n-p, \alpha}$, o modelo em questão será rejeitado e caso seja menor ou igual aceito (CORDEIRO; PAULA, 1989).

A expressão do desvio (ou função desvio) para o MLG binomial para dados na forma de proporção é determinado como segue

Para o modelo saturado, tem $\tilde{\theta}_i = \log \frac{y_i}{1 - y_i}$ e $b(\tilde{\theta}_i) = -\log(1 - y_i)$. Também $a(\phi) = 1/m_i$, assim $\phi = 1$ e $w_i = m_i$. Similarmente para qualquer outro modelo com $p < n$ parâmetros, tem-se $\hat{\theta}_i = \log \frac{\hat{\pi}_i}{1 - \hat{\pi}_i}$ e $b(\hat{\theta}_i) = \log[1 + \exp(\theta_i)] = -\log(1 - \pi_i)$ e de acordo com 17, tem-se

$$\begin{aligned} D_p &= 2 \sum_i m_i \left[y_i \left(\log \frac{y_i}{1 - y_i} - \log \frac{\hat{\pi}_i}{1 - \hat{\pi}_i} \right) + \log(1 - y_i) - \log(1 - \hat{\pi}_i) \right] \\ &= 2 \sum_i m_i y_i [\log y_i - \log(1 - y_i) - \log \hat{\mu}_i + \log(1 - \hat{\mu}_i)] + \log(1 - y_i) - \log(1 - \hat{\mu}_i) \end{aligned}$$

$$\begin{aligned}
&= 2 \sum_i m_i [y_i \log y_i - y_i \log(1 - y_i) - y \log \hat{\mu}_i + y_i \log(1 - \hat{\mu}_i) + \log(1 - y_i) - \log(1 - \hat{\mu}_i)] \\
&= 2 \sum_i m_i [y_i(\log y_i - \log \hat{\mu}_i) + \log(1 - y_i)(-y + 1) + \log(1 - \hat{\mu}_i)(y_i - 1)] \\
&= 2 \sum_i m_i \left[y \log \frac{y_i}{\hat{\mu}_i} + (1 - y_i) \log(1 - y_i) - (1 - y_i) \log(1 - \hat{\mu}_i) \right] \\
&= 2 \sum_i m_i \left[y_i \log \frac{y_i}{\hat{\mu}_i} + (1 - y_i) \log \frac{1 - y_i}{1 - \hat{\mu}_i} \right] \tag{53}
\end{aligned}$$

Sendo $m_i y_i$ o número de sucessos e $(m_i - m_i y_i)$ o número de insucessos, $i = 1, 2, \dots, n$.

3.10 Componentes do desvio do MLG binomial

Para o MLG binomial, considerando dados na forma de proporção, os componentes do desvio d_i , tem expressão dada por,

$$d_i = -2m_i \left[y_i \log \frac{y_i}{\hat{\mu}_i} + (1 - y_i) \log \frac{1 - y_i}{1 - \hat{\mu}_i} \right], \text{ se } y_i \neq 0, 1$$

$$d_i = 2m_i [-y_i \log \hat{\mu}_i - (1 - y_i) \log(1 - \hat{\mu}_i)], \text{ se } y_i \in \{0, 1\}.$$

Assim, os resíduos componentes do desvio, que são definidos como iguais às raízes quadradas dos componentes do desvio, tem expressão da forma,

$$r_i^D = \left[-2m_i \left(y_i \log \frac{y_i}{\hat{\mu}_i} + (1 - y_i) \log \frac{1 - y_i}{1 - \hat{\mu}_i} \right) \right]^{1/2}, \text{ se } y_i \neq 0, 1$$

$$r_i^D = [2m_i (y_i \log \hat{\mu}_i + (1 - y_i) \log(1 - \hat{\mu}_i))]^{1/2}, \text{ se } y_i \in \{0, 1\}.$$

A análise confirmatória da qualidade do ajuste foi verificada pelo gráfico de probabilidade normal padrão com envelope de simulação, usando os resíduos componentes do desvio (PAULA, 2013). O envelope de simulação é obtido por meio de simulações de amostras de acordo com a distribuição pressuposta, obtendo-se para cada ponto do gráfico um valor como um limite inferior e um valor como limite superior. Conclui-se pela adequação do modelo probabilístico se todos os pontos plotados estiverem predominantemente dentro do envelope com possível exceção de alguns pontos. Como se vê, a interpretação ainda carrega um certo grau de subjetividade, mas, em geral, o envelope auxilia bastante na interpretação.

3.11 Análise de desvio - ANODEV

Para a seleção dos modelos, foi utilizada a análise de desvio - ANODEV, apresentado na Tabela 1. Cujas estatísticas de teste considera a diferença dos desvios $(D_q - D_p)$, com $p - q$ número de graus de liberdade entre os modelos M_{D_q} e M_{D_p} ($q < p$) comparados. Assim, se $(D_q - D_p) > \chi^2_{p-q, \alpha}$, os efeitos dos termos que estão em M_{D_p} e não estão em M_{D_q} são significativos.

Uma outra estatística utilizada, para o caso em que usou-se peso na variável binomial, foi a estatística,

$$F = \frac{(D_q - D_p)/(p - q)}{\hat{\phi}}$$

Assim, se $F > F_{p-q, n-m}$ os efeitos dos termos que estão em M_{D_p} e não estão em M_{D_q} são significativos, sendo $\hat{\phi}$ estimado a partir do modelo M_{D_p} (modelo com maior número de parâmetros).

3.12 Viabilidade dos ovos de *Aedes aegypti*

A viabilidade dos ovos de *Aedes aegypti* foi predita usando a ideia dos modelos de dose-resposta. Estes modelos, visam não só a predição da probabilidade de sucesso $\pi(x)$ para uma dosagem específica x , mas também a determinação da dosagem necessária para se atingir uma probabilidade de sucesso p . Essa dosagem é chamada de dose letal. A notação usual para uma dose letal de $100p\%$ é dada por DL_{100} . Logo, $p = \pi(\beta_0 + \beta_1 DL_{100p})$, com $0 < p < 1$, sendo a mais comum em toxicologia a dose mediana (DL_{50}) (PAULA, 2013).

Modelos de dose-resposta são empregados em várias áreas do conhecimento, em que a dose pode ser a idade, o peso, a resistência de um certo material (PAULA, 2013). Trabalho usando a ideia desta metodologia, pode ser encontrado em (MILECER; SZCZOTKA, 1966 apud PAULA, 2013).

Deste modo, considerando o ajuste do MLG binomial com preditor linear $\eta_i = \beta_0 + \beta_1 x_i$, o estimador de máxima verossimilhança $\widehat{DL}_{50}$ do parâmetro DL_{50} , fica dado por:

Para função de ligação logística,

$$\widehat{DL}_{50} = d(\hat{\beta}) = \frac{1}{\beta_1} \left[\log \frac{\pi}{1 - \pi} - \beta_0 \right] \quad (54)$$

que para $\pi=0,5$, tem-se

$$\widehat{DL}_{50} = \frac{\hat{\beta}_0}{-\hat{\beta}_1} \quad (55)$$

Para função de ligação probito,

$$\widehat{DL}_{50} = d(\hat{\beta}) = \frac{1}{\beta_1} [\Phi^{-1}(\pi) - \beta_0] \quad (56)$$

que para $\pi=0,5$, tem-se

$$\widehat{DL}_{50} = \frac{\hat{\beta}_0}{-\hat{\beta}_1}$$

em que $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T$ e o estimador de máxima verossimilhança de $\beta = (\beta_0, \beta_1)^T$.

Devido a simetria dos MLGs com função de ligação logística ou função de ligação probito, $\widehat{DL}_{50}$ estima o tempo mediano que os ovos de *Aedes aegypti* permanecem em condições de eclodir e produzir larvas viáveis após permanecerem armazenados por determinado período (viabilidade).

3.13 Intervalo de confiança para DL_{50}

Os métodos comumente usados para a construção de intervalos de confiança para doses efetivas são: Metodo Delta, o metodo de Fieller e o metodo da razão de verossimilhança. Uma descrição bem mais detalhada sobre construção de intervalos de confiança pode ser encontrada em (COLLETT, 2003).

3.13.1 Método Delta

Segundo Paula (2013), o método delta calcula a variância assintótica de uma função escalar $g(\beta)$ de um vetor β de dimensão p , de parâmetros desconhecidos, quando a matriz de covariância de $\hat{\beta}$ é conhecida. É baseado na expansão de Taylor até a primeira ordem e supõe-se que, segundo condições gerais de regularidade, a distribuição assintótica do EMV $\hat{\beta}$ é $N_p(\beta, \mathbf{V})$, sendo $\mathbf{V} = \hat{\mathbf{K}}^{-1}$ a matriz de variâncias e covariâncias dos estimadores dos parametros (obtida pela inversa da matriz de informação).

Assim, um intervalo de $100(1 - \alpha)\%$ de confiança para DL_{50} , em grandes amostras, considerando-se a função de ligação logística é dado por,

$$\widehat{DL}_{50} \pm z_{1-\frac{\alpha}{2}} \sqrt{\widehat{Var}_A[d(\hat{\beta})]} \quad (57)$$

Uma vez conhecida a covariância assintótica entre β 's, pode-se calcular matriz de variância assintótica de $\widehat{DL}_{50}$. Esta pode ser determinada após uma aproximação de primeira ordem

por série de Taylor de $d(\hat{\beta})$ em torno de β , cuja a expressão é da forma

$$Var_A[\widehat{DL}_{50}] = D(\beta)^T (X^T V X)^{-1} D(\beta),$$

em que

$$D(\beta) = \partial d(\beta) / \partial \beta = \left[\frac{-1}{\beta_1}, \frac{1}{\beta_1^2} \{ \beta_0 - \log(\frac{\pi}{1-\pi}) \} \right]^T$$

e $(X^T \hat{V} X)^{-1}$ contém as variâncias e covariâncias assintóticas estimadas de $\hat{\beta}_0$ e $\hat{\beta}_1$ e $z_{1-\alpha/2}$, valor da distribuição normal padrão

3.14 Casos especiais do MLG binomial

Cordeiro e Demétrio (2011) apresentam casos especiais do modelo binomial, obtidos a partir da definição de uma distribuição de tolerância. Supondo que a tolerância tem distribuição normal, o modelo correspondente $\pi_i = \Phi(\alpha + \beta x_i)$ é denominado probito. Se se supõe que a tolerância tem distribuição logística, o modelo $\pi_i = \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}$ é denominado logístico. A dedução destes é a apresentada a seguir.

3.14.1 Modelo logístico

Assumindo que a variável latente T (tolerância ao estímulo) tem distribuição logística com parâmetros μ pertencente aos reais e $\tau > 0$, e função densidade de probabilidade expressa por

$$f(t; \mu, \tau) = \frac{1}{\tau} \frac{\exp(\frac{\tau - \mu}{\tau})}{[1 + \exp(\frac{\tau - \mu}{\tau})]^2} \quad (58)$$

com média $E(T) = \mu$ e variância $\sigma^2 = Var(T) = \pi^2 \tau^3 / 3$. Na expressão em (58), fazendo $\alpha = -\mu/\tau$ e $\beta = 1/\tau$ tem-se,

$$f(t; \mu, \tau) = \frac{\beta \exp(\alpha + \beta t)}{[1 + \exp(\alpha + \beta t)]^2} \quad (59)$$

sendo $\pi_i = P(T \leq x_i) = F(x_i)$ obtem-se,

$$\begin{aligned} \pi_i = F(x_i) &= \int_{-\infty}^{x_i} \frac{\beta \exp(\alpha + \beta t)}{[1 + \exp(\alpha + \beta t)]^2} dt = - \lim_{a \rightarrow -\infty} \left[\frac{1}{1 + \exp(\alpha + \beta t)} \right]_a^{x_i} \\ &= - \frac{1}{1 + \exp(\alpha + \beta x_i)} + 1 \end{aligned}$$

assim

$$\pi_i = \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}$$

é uma função não linear em um conjunto linear de parâmetros, sendo linearizada por

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \alpha + \beta x_i.$$

3.14.2 Modelo probito

Assume-se que a variável latente T , tem distribuição normal com média μ e variância σ^2 . Isto é,

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

deste modo, a variável

$$Z = \frac{T - \mu}{\sigma} \sim N(0, 1)$$

Então

$$\pi_i = P(T \leq x_i) = P\left(Z \leq -\frac{\mu}{\sigma} + \frac{1}{\sigma}x_i\right) = P(Z \leq \alpha + \beta x_i)$$

em que, $\alpha = \mu/\sigma$ e $\beta = 1/\sigma$. Logo

$$\pi = \Phi(\alpha + \beta x_i)$$

que é uma função não linear em um conjunto linear de parâmetros, sendo que $\Phi(\cdot)$ representa a função de distribuição normal padrão. A linearização é dada por

$$\text{probit}(\pi_i) = \Phi_i(\pi_i) = \alpha + \beta x_i.$$

4 RESULTADOS E DISCUSSÃO

A seguir são apresentados e discutidos os resultados obtidos a partir da análise do conjunto de dados (Tabela 2). Todos foram obtidos a partir do software estatístico R (R DEVELOPMENT CORE TEAM, 2013) e os comandos usados encontram-se no anexo.

Inicialmente é apresentada na Figura 3, a relação idade dos ovos de *Aedes aegypti* versus proporções de eclosões de larvas. Observa-se para o recipiente copo plástico, nos cinco primeiros períodos da idade uma proporção de eclosões com maior destaque em relação aos recipientes envelope de papel e saco plástico. Ainda, uma acentuada baixa na proporção de eclosão para o período de 19 dias do recipiente saco plástico (observação 19), aparentemente um ponto atípico. Segundo Demétrio e Zocchi (2011), pontos atípicos (falhas isoladas) podem surgir devido a erros grosseiros na variável resposta ou nas variáveis explanatórias, por medidas erradas, ou registro da observação, ou ainda erros de transcrição, e outros. Em conversa com os pesquisadores, eles justificaram, o que pode ter ocorrido para explicar a acentuada baixa na proporção de eclosão deste caso é que havia uma quantidade de ovos inviáveis, e que passou despercebido quando foi realizada a contagem. Observa-se também um comportamento sigmoidal.

4.1 Modelos ajustados

Para os modelos lineares generalizados ajustados, caso em que considerou-se a função de ligação probito e preditores lineares em (35) e (36) os valores dos desvios residuais são 263,60 (com 42 graus de liberdade) e 380,31 (com 44 graus de liberdade) respectivamente, mostrando evidências da falta de ajuste . Ainda, na Figura 4, tem-se os gráficos das proporções de larvas de *Aedes aegypti* versus a idade em que os ovos foram tratados e as curvas ajustadas.

Para os ajustes, em que considerou-se a função de ligação logística e preditores lineares em (35) e (36), os valores dos desvios residuais são 244,31 (com 42 graus de liberdade) e 378,08 (com 44 graus de liberdade), mostrando evidências da falta de ajuste para os dois modelos. Na figura 5, tem-se os gráficos das proporções de larvas de *Aedes aegypti* versus a idade em que os ovos foram tratados e as curvas ajustadas. Apenas como ilustração são mostradas as Figuras 4 e 5.

Figura 4 – Gráficos das proporções de larvas de *Aedes aegypti* versus a idade em que os ovos foram tratados e as curvas ajustadas pelo MLG com ligação probito, utilizando-se os preditores lineares retas concorrentes (a) e retas paralelas (b)

Figura 5 – Gráficos das proporções de larvas do *Aedes aegypti* versus a idade em que os ovos foram tratados e as curvas ajustadas pelo MLG com ligação logística, utilizando-se os preditores lineares retas concorrentes (a) e retas paralelas (b)

Dado que se constatou a falta de ajustes dos modelos iniciais, novos ajustes foram realizados, em que optou-se pela inclusão de um peso ($w = 1/\hat{\phi}$) na variável binomial mediante o argumento *weights* da função *glm* do programa R. Assim, os valores dos desvios residuais, para os novos ajustes em que considerou-se a função de ligação probito e preditores lineares em (35) e (36) são 37,788 (com 42 graus de liberdade) e 40,817 (com 44 graus de liberdade), com respectivos valores do nível descritivo ($p=0,6563$ e $p=0,6088$), mostrando evidências que os modelos com retas concorrentes e retas paralelas ajustam-se bem aos dados.

Embora os desvios residuais tenham apresentado evidências de ajustes adequados, o gráfico de probabilidade normal com envelopes simulados, Figura 6a, apresenta vários pontos fora do envelope e a Figura 6b apresenta dois pontos com mais destaque (19 e 26), o que evidencia a falta de ajuste dos modelos. Ainda, na Figura 7, tem-se os gráficos das proporções de larvas versus a idade em que os ovos foram tratados e as curvas ajustadas, sendo $wc = 1/6,9757$ e $wp = 1/9,3175$ os pesos adicionados na variável binomial para os ajustes com retas concorrentes e paralelas. As Tabelas 3 e 4 apresentam a análise de desvios para ambos os modelos.

Figura 6 – Gráficos de probabilidade normal com envelopes simulados referentes ao MLG binomial com função de ligação probito e preditores lineares retas concorrentes (a) e retas paralelas (b)

Figura 7 – Gráficos das proporções de larvas de *Aedes aegypti* versus a idade em que os ovos foram tratados e as curvas ajustadas pelo MLG com inclusão de peso na variável binomial com ligação probito e preditores retas concorrentes (a) e retas paralelas (b)

Tabela 3 – Análise de desvio para o MLG binomial com função de ligação probito e preditores lineares retas paralelas

Fonte de variação	g.l	Desvios	F	Valor de p
Recipiente	2	10,19	5,095	0,010
Regressão	1	582,3	582,3	<0,001
Resíduo	44	40,817		
Total	47	633,312		

Tabela 4 – Análise de desvio para o MLG binomial com função de ligação probito e preditores lineares retas concorrentes

Fonte de variação	g.l	Desvios	F	Valor de p
Recipiente	2	13,61	6,805	0,003
Regressão	3	794,59	264,87	<0,001
Resíduo	42	37,788		
Total	47	846,01		

Para o caso em que se utilizou a função de ligação logística e preditores lineares (35) e (36), os valores dos desvios residuais são 39,25 (com 42 graus de liberdade) e 40,14 (com 44 graus de liberdade) com respectivos níveis descritivos ($p = 0,592$ e $p = 0,637$), mostrando evidências que os modelos com retas concorrentes e retas paralelas ajustam-se bem aos dados.

Na Figura 8, tem-se os gráficos de probabilidade normal com envelopes simulados, mostrando evidências da falta de ajuste de ambos os modelos e, consequentemente não confirmando a análise dos desvios residuais. Ainda, na Figura 9, tem-se os gráficos das proporções de larvas *versus* a idade em que os ovos foram tratados e as curvas ajustadas, sendo respectivamente $wc = 1/6,2245$ e $wp = 1/9,4193$ os pesos adicionados na variável binomial para os ajustes com retas concorrentes e retas paralelas. As Tabelas 5 e 6 apresentam a análise de desvios para os dois modelos.

Figura 8 – Gráficos de probabilidade normal com envelopes simulados referentes ao MLG binomial com função de ligação logística e preditores lineares retas concorrentes (a) e retas paralelas (b)

Figura 9 – Gráficos das proporções de larvas de *Aedes aegypti* versus a idade em que os ovos foram tratados e as curvas ajustadas, considerando-se a ligação logística e os preditores lineares retas concorrentes (a) e retas paralelas (b)

Tabela 5 – Análise de desvio para o MLG binomial com função de ligação logística e preditores lineares retas paralelas

Fonte de variação	g.l	Desvios	F	valor de p
Recipiente	2	10,08	5,04	0,011
Regressão	1	576,27	576,25	<0,001
Resíduo	44	40,14		
Total	47	626,48		

Tabela 6 – Análise de desvio para o MLG binomial com função de ligação logística e preditores lineares retas concorrentes

Fonte de variação	g.l	Desvios	F	valor de p
Recipiente	2	15,26	7,628	0,001
Regressão	3	893,58	297,83	<0,001
Resíduo	42	39,25		
Total	47	948,09		

Paula (2013) e Fox (2002) discutem como possível alternativa para contornar o problema da falta de ajuste ocasionada pela presença de pontos aberrantes/influentes a retirada destes do conjunto de dados. Discutem problemas em que sugerem a retirada de uma observação ou duas observações com a finalidade de melhorar a qualidade do ajuste do modelo. Segundo Terra (2009), observações influentes podem acarretar mudanças nas estimativas dos parâmetros e até mesmo causar mudanças inferenciais.

Reajustando o modelo linear generalizado, com função de ligação logística sem duas observações que apareciam com mais destaques (observação de número 19 e de número 26), destacadas nas Figuras 6 e 8, gráficos de probabilidade normal com envelopes simulados.

A Tabela 7 apresenta os valores dos desvios residuais e seus respectivos números de graus de liberdade (g.l). Verifica-se que existem evidências de que os modelos com retas concorrentes e com retas paralelas ajustam-se bem aos dados. Nas Tabelas 8 e 9, tem-se a análise de desvios para os modelos. Observa-se que, em ambas existem

evidências para o efeito significativo do fator recipiente e para o efeito significativo da regressão.

A Figura 10a, gráfico de probabilidade normal com envelopes simulados, apresenta pelo menos quatro pontos fora dos limites e os resíduos distribuídos de forma um tanto irregular dentro do envelope, o que evidencia a falta de ajuste do modelo com retas concorrentes. Já pela Figura 10b, gráfico de probabilidade normal com envelopes simulados, nota-se todos os pontos dentro do envelope e bem melhores distribuídos, evidenciando que o modelo com retas paralelas ajusta-se bem aos dados.

Na Figura 11, tem-se os gráficos das proporções de larvas *versus* a idade em que os ovos foram tratados e as curvas ajustadas, sendo $w_c = 1/4, 10$ e $w_p = 1/6, 2685$ os pesos considerados nos respectivos ajustes. Resultados semelhantes de diagnóstico por envelopes podem ser encontrados em (PAULA, 2013). O autor discute ainda resultados em que uma observação ou duas observações aparecem fora do envelope porém, muito próximas aos limites e atesta um bom ajuste do modelo investigado.

Figura 10 – Gráficos de probabilidade normal com envelopes simulados referentes ao MLG binomial com função de ligação logística e preditores lineares retas concorrentes (a) e retas paralelas (b)

Figura 11 – Gráficos das proporções de larvas do *Aedes aegypti* versus a idade em que os ovos foram tratados e as curvas ajustadas pelo MLG com ligação logística, utilizando-se como preditores lineares retas concorrentes (a) e retas paralelas (b)

Tabela 7 – Desvios residuais do MLG binomial com função de ligação logística

Modelo	g.l	Desvios	Valor de p
$\eta = \alpha_j + \beta_j x$	40	40,424	0,4515
$\eta = \alpha_j + \beta x$	42	43,59	0.4032

Tabela 8 – Análise de desvio para o MLG binomial com função de ligação logística e preditor linear retas concorrentes

Fonte de variação	g.l	Desvios	F	valor de p
Recipiente	2	23,23	11,63	<0,001
Regressão	3	1373,29	457,67	<0,001
Resíduo	40	40,424		
Total	45	1436,945		

Tabela 9 – Análise de desvio para o MLG binomial com função de ligação logística e preditor linear retas paralelas

Fonte de variação	G.l	Desvios	F	Valor de p
Recipiente	2	15,19	7,597	0,002
Regressão	1	881,1	881,1	< 0,001
Resíduo	42	43,60		
Total	45	939,89		

A partir dos resultados obtidos, considerando-se diversos ajustes, verifica-se que o MLG binomial com ligação logística e preditor linear retas paralelas, com peso $w = 1/6,2685$ na variável binomial em que eliminou-se duas observações, é o modelo que melhor explica a relação proporção de eclosões e período de armazenamento. Sendo considerados o desvio residual e o gráfico de probabilidade normal padrão com envelope de simulação como medida da qualidade do ajuste e análise de desvio (ANODEV) como critério de seleção de modelos.

As Tabelas 10 e 11 apresentam as estimativas dos parâmetros bem como os erros padrão para o modelo com todas as observações (Figura 9b) e para o modelo em que optou-se pela retirada das observações 19 e 26 (Figura 11b). Verifica-se, que as estimativas dos parâmetros são bem próximas em ambos os casos, os erros padrão apresentam uma pequena redução para o modelo em que eliminou-se duas observações e os resultados inferenciais não são alterados.

Tabela 10 – Estimativas dos parâmetros do MLG binomial com função de ligação logística e preditor linear retas paralelas considerando todas as observações

Parâmetros	Estimativas	E.Padrão	Valor de t	Valor de p
β_{01}	4,291	0,316	13,56	<0,001
β_{02}	3,398	0,2779	12,23	<0,001
β_{03}	3,522	0,2816	12,51	<0,001
β_1	-0,041	0,003	-15,89	<0,001

Tabela 11 – Estimativas dos parâmetros do MLG binomial com função de ligação logística e preditor linear retas paralelas em que eliminou-se duas observações

Parâmetros	Estimativas	E.Padrão	Valor de t	Valor de p
β_{01}	4,4612	0,2712	16,45	<0,001
β_{02}	3,4577	0,2523	13,70	<0,001
β_{03}	3,6731	0,2413	15,22	<0,001
β_1	-0,0423	0,002	-19,07	<0,001

A partir do MLG binomial com ligação logística selecionado, obtem-se, respectivamente para copo plástico, envelope de papel e saco plástico as equações,

$$\log \left(\frac{\hat{\pi}_1(x_i)}{1 - \hat{\pi}_1(x_i)} \right) = 4,461 - 0.042x_i \quad (60)$$

$$\log \left(\frac{\hat{\pi}_2(x_i)}{1 - \hat{\pi}_2(x_i)} \right) = 3,458 - 0.042x_i \quad (61)$$

$$\log \left(\frac{\hat{\pi}_3(x_i)}{1 - \hat{\pi}_3(x_i)} \right) = 3,673 - 0,042x_i \quad (62)$$

logo, a viabilidade mediana dos ovos de *Aedes Aegypti* para copo plástico, saco plástico e envelope de papel juntamente com seus respectivos intervalos de confiança são:

$$\text{Copo plástico: } \widehat{DL}_{50} = \frac{4,461}{0,042} = 106 \text{ (98 ; 114) dias;}$$

$$\text{Saco plástico: } \widehat{DL}_{50} = \frac{3,673}{0,042} = 87 \text{ (80 ; 93) dias;}$$

$$\text{Envelope de papel: } \widehat{DL}_{50} = \frac{3,458}{0,042} = 83 \text{ (73 ; 93) dias.}$$

Na Figura (11), verifica-se que as curvas ajustadas para os recipientes saco plástico e envelope de papel estão bem próximas e, portanto, é interessante testar se elas diferem estatisticamente. Assim, um novo modelo foi usado, considerando-se os dados dos recipientes envelope de papel e saco plástico como se fosse um único recipiente. Cordeiro e Demétrio (2011) apresentam procedimento idêntico num estudo com gemas de galhos de três variedades de macieiras, com o propósito de verificar se as variedades diferem estatisticamente.

O valor do desvio residual para o novo modelo, considerando a função de ligação logística e preditor linear (36) é 45,31 (com 43 graus de liberdade) com nível descritivo ($p = 0,3756$), evidencia que o modelo com retas paralelas ajusta-se bem aos dados. A diferença de desvios residuais entre o novo modelo e o modelo selecionado (Figura 10b) é (4,5534) com ($F=0,7264$ e $p = 0.3989$), indicando existir evidências de que os recipientes envelope de papel e saco plástico comportam-se de forma semelhante. Na Tabela 12, tem-se a análise de desvios. Nesta, observa-se que existem evidências para o efeito do fator recipiente e para o efeito da regressão.

Tabela 12 – Análise de desvio para o novo modelo em que utilizou-se a função de ligação logística e preditor linear retas paralelas

Fonte de variação	G.l	Desvios	F	Valor de p
Recipiente	1	13,55	13,55	$< 0,001$
Regressão	1	901,99	901,99	$< 0,001$
Resíduo	43	45,31		
Total	45	960,87		

Um dos fatores que pode explicar a maior viabilidade (106 dias) ocorrida no recipiente copo é que estes recipientes ficavam abertos, proporcionando aos ovos um contato maior com o ar. Já os ovos armazenados nos recipientes envelopes de papel e sacos plástico ficavam fechados, menos ar. Uma outra leitura é que realmente exista o efeito do recipiente copo sobre a viabilidade.

Pinheiro (2005) compara a viabilidade média dos ovos de *Aedes aegypti* entre os recipientes copo plástico, saco plástico e envelope de papel, usando a abordagem de ANOVA e teste de Tukey. Constata não haver diferenças significativas entre saco plástico e envelope de papel, ambos diferindo do recipiente copo plástico que apresentou a maior viabilidade, corroborando nesse sentido os resultados aqui apresentados. Ainda, estima uma viabilidade média de 98 dias, não fazendo referências para qual tipo de recipiente.

5 CONCLUSÕES

A utilização do modelo linear generalizado, com peso $w = 1/6,2685$ na variável binomial, mediante o argumento *weights* da função *glm* do programa R, para fins de diminuição da variação presente, apresentou-se como uma boa alternativa de modelagem.

No presente estudo, o modelo linear generalizado binomial, com função de ligação logística e preditor linear retas paralelas, ajustou-se bem aos dados, possibilitando estimar satisfatoriamente o parâmetro DL_{50} , sendo esta a maior contribuição deste trabalho. Entretanto, o bom resultado do modelo só foi possível, mediante a retirada (eliminação) de duas observações do conjunto de dados. Observações estas que, quando retiradas, não alteraram as estimativas dos parâmetros e as inferências. Ainda, a análise da qualidade do ajuste deu-se mediante o gráfico de probabilidade normal padrão com envelope de simulação sobre os resíduos componentes do desvio.

Portanto, o modelo linear generalizado para dados na forma de proporção com peso na variável binomial descreve bem o comportamento da viabilidade dos ovos de *Aedes aegypti* criados em condições laboratoriais e mantidos em diferentes tipos de armazenamentos. A viabilidade estimada para os ovos armazenados nos recipientes copo plástico, saco plástico e envelope de papel são 106, 87 e 83 dias respectivamente. Também foi possível verificar diferenças significativas nas estimativas medianas da viabilidade dos ovos de *Aedes aegypti*, sendo que para o recipiente copo plástico, obteve-se a maior viabilidade.

Acrescenta-se como proposta de continuidade do tema abordado nesta dissertação o desenvolvimento dos seguintes trabalhos futuros: Predição da viabilidade para períodos de armazenamentos igualmente distantes estudados ao longo de 450 dias, assim como a comparação dos parâmetros dos modelos ajustados: modelo binomial com peso na variável resposta *versus* modelo beta binomial.

REFERÊNCIAS

- AGRESTI, A. **Categorical data analysis**. 2nd ed. Hoboken: John Wiley, 2002. 710 p.
- BOX, G.E.P.; COX, D.R. An analysis of transformations. **Journal of the Royal Statistical Society**. Series B, London, v. 26, n. 2, p. 211-252, 1964.
- BRAGA, I.A.; VALLE, D. Aedes aegypti: histórico do controle no Brasil. **Epidemiologia e Serviços de Saúde**, Brasília, v. 16, n. 2, p. 113-118, jun. 2007.
- COLLETT, D. **Modelling binary data**. 2nd ed. Boca Raton: Chapman & Hall; CRC Press, 2003. 387 p. (Texts in Statistical Science Series).
- CORDEIRO, G.M.; DEMÉTRIO, C.G.B. **Modelos lineares generalizados e extensões**. 2011. Disponível em: <[http://www4.esalq.usp.br/departamentos/lce/arquivos/aulas/2010/LCE5868 /livro.pdf](http://www4.esalq.usp.br/departamentos/lce/arquivos/aulas/2010/LCE5868/livro.pdf)>. Acesso em: 23 jan. 2013.
- CORDEIRO, G.M.; DEMÉTRIO, C.G.B. **Modelos lineares generalizados**. Santa Maria: ESALQ, 2007. 161 p.
- CORDEIRO, G. M.; LIMA NETO, E.A. **Modelos paramétricos**. São Paulo: ABE, 2004. 246 p.
- CORDEIRO, G.M.; PAULA, G.A. **Modelos de regressão para análise de dados univariados**. 1989. Disponível em: <http://www.impa.br/opencms/pt/biblioteca/cbm/17CBM/17_CBM_89_04.pdf>. Acesso em: 06 nov. 2013.
- CRAWLEY, M.J. **The R book**. Wiltshire: John Wiley, 2007. 942 p.
- DEMÉTRIO, C.G.B.; ZOCCHI, S.S. **Modelos de regressão**. Piracicaba: ESALQ, 2011. 216 p.
- DOBSON, A.J. **An introduction to generalized linear models**. 2nd ed. Boca Raton: Chapman & Hall; CRC Press, 2002. 221 p. (Texts in Statistical Science Series).
- FORANTTINI, O.P. **Culicidologia médica**. São Paulo: EDUSP, 2002. v. 2, 864 p.
- FOX, J. **An R and S-PLUS companion to applied regression**. Thousand Oaks: Sage Publications, 2002. 312 p.
- FUNDAÇÃO NACIONAL DE SAÚDE. **Dengue instruções para pessoal de combate ao vetor**. Brasília, 2001. 84 p.
- JACHMAN, S. **Generalized linear model**. Stanford University, 2011. Disponível em: <<http://jackman.stanford.edu/papers/glm.pdf>>. Acesso em: 12 jun.2013.

LEE, Y.; NELDER, J.A. Joint modeling of mean and dispersion. **Technometrics**, Arlington, v. 40, n. 2, p. 168-175. 1998.

McCULLAGH, P.; NELDER, J.A. **Generalized linear models**. 2nd ed. London; New York: Chapman and Hall, 1989. 511 p.

MENÉNDEZ FERNÁNDEZ, J.A. **Un modelo logístico com sobredispersión y ajuste quasibinomial**. 2013. Disponível em: <<http://www.eio.uva.es/~josan/mlgweb/logisticos/grainbeetles.pdf>>. Acesso em: 24 maio 2013.

MYERS, R.H.; MONTGOMERY, D.C. A Tutorial on generalized linear models. **Journal of Quality Technology**, Milwaukee, v. 29, n. 3, p. 274-291, 1997.

MONATH, P.T. Dengue: the risk to developed and developing countries. **Proceedings of the National Academy of Sciences of the USA**, Washington, v. 91, p. 2395-2400, Mar. 1994.

NELDER, J.A.; LEE, Y. Generalized linear models for the analysis of Taguchi-TyPe experiments. **Applied Stochastic Models and Data Analysis**, Hoboken, v. 7, n. 1, p. 107-120, 1991.

NELDER, J.A.; WEDDERDURN, R.W.M. Generalized linear models. **Journal of the Royal Statistical. Series A: General Statistics**, London, v. 135, p. 370-384, 1972.

PAULA, G.A. **Modelos de regressão com apoio computacional**. Disponível em: <http://www.ime.unicamp.br/~cnaber/Livro_MLG.pdf>. Acesso em: 25 jan. 2013.

PAULINO, C.D.; SINGER, J.M. **Análise de dados categorizados**. São Paulo: Edgard Blucher, 2006. 629 p.

PINHEIRO, V.C.S. **Controle do dengue em Manaus AM: estudo da positividade/produtividade larvária, viabilidade dos ovos e investigação por RTPCR dos sorotipos do vírus dengue em *Aedes aegypti* (Linnaeus, 1762) (Diptera, Culicidae)**. 2005. 167 p. Tese (Doutorado em Ciências Biológicas) - Instituto Nacional de Pesquisas da Amazônia; Universidade Federal do Amazonas, Manaus, 2005.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2013. Disponível em: <<http://www.R-project.org/>>. Acesso em: 18 nov. 2013.

RAO, C.R ; WU, Y. Linear model selection by cross-validation, **Journal Statistical Planning and Inference**, Toronto, v. 128, n. 1, p. 231-240, 2005.

SAMPAIO, S. C. **A longevidade do *Aedes aegypti* durante a estação chuvosa no município de Fortaleza**. 2010. 114 p. Dissertação (Mestrado em Saúde Pública) - Universidade Estadual do Ceará, Fortaleza, 2010.

SANTOS, M.A.V.M. ***Aedes aegypti* (DIPTERA: CULICIDAE): estudos populacionais e estratégias integradas para controle vetorial em municípios da região metropolitana do Recife, no período de 2001 a 2007**. 2008. 218 p. Tese (Doutorado em Saúde Pública) - Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2008.

SCARPASSA, V.M.; TADEI, W.P. Biologia de anofelinos amazônicos. XIII. Estudo do ciclo biológico de *Anopheles nuneztovari*. **Acta Amazônica**, Manaus, v. 20, p. 95-118, 1990.

SILVA, H.H.G. da; SILVA, I.G. da. Influência do Período de Quiescência dos Ovos Sobre o Ciclo de Vida de *Aedes aegypti* (Linnaeus, 1762) (Diptera, Culicidae) em Condições de Laboratório. **Revista da Sociedade Brasileira de Medicina Tropical**, Goiás, v. 32, n. 4, p. 349-355, 1999.

SILVA, I.G.; CAMARGO, M.F.; ELIAS, M.; ELIAS, C.N. Ciclo Evolutivo de *Aedes* (Stegomyia) *aegypti* (Linnaeus, 1762) (Diptera: Culicidae). **Revista de Patologia Tropical**, Goiás, v. 22, n. 1, p. 43-48. 1993.

TAUIL, P.L. Urbanização e ecologia do dengue. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 17, p. 99-102, 2001. Suplemento.

TERRA, L.C. **Modelos lineares generalizados simétricos heteroscedásticos**. 2009. 48 p. Dissertação (Mestrado em Estatística) - Universidade Federal de Pernambuco, Recife, 2009.

URBANO, M.R. **Modelagem de mortalidade natural e superdispersão em dados entomológicos**. 2012. 117 p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2012.

VIEIRA, A.F.C. **Análise da média e dispersão em experimentos fatoriais não replicados para a otimização de processos industriais**. 2004. 204 p. Tese (Doutorado em Engenharia de Produção) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2004.

ANEXOS

Código em R para o ajuste dos modelos

```
rm(list=ls(all=TRUE))
#=====
#Importação dos dados do arquivo finalmod
#=====
dados<-read.table("finalmod.txt",h=T)
dados
attach(dados)
names(dados)
  (rec<-dados[,1])#tipos de recipientes
  (ida<-dados[,2])# idade dos ovos/ periodo de armazenamento
  (total<-dados[,3])# total de ovos
  (eclosao<-dados[,4])# eclosão de larvas
  (prop<-eclosao/total )# proporção de larvas
  (dado<-data.frame(rec,ida,ide,total,eclosao,prop))
#=====
# Diagrama de dispersão
#=====
resp<-cbind(eclosao,total-eclosao)
pfc<-eclosao/total
pfc<-split(prop,rec)
dc<-split(ida,rec)
plot(idade,proporção,type="n",ylab="Proporção de larvas")
points(dc[[1]],pfc[[1]],pch=16,col="red")
points(dc[[2]],pfc[[2]],pch=14,col="red")
points(dc[[3]],pfc[[3]],pch=17,col="blue")

#=====
#ca1- Ajuste do MLG binomial com função de ligação logística.
#=====
mod1<-glm(resp~rec+ida-1,family=binomial(link="logit")#rtas paralelas
mod2<-glm(resp~rec/ida-1,family=binomial(link="logit")#rtas concorrentes

#=====
caso 1 - ajuste de MLG binomial com função de ligação probito
#=====
mod3<-glm(resp~rec+ida-1,family=binomial(link="probit")#rtas paralelas
mod4<-glm(resp~rec/ida-1,family=binomial(link="probit")#rtas concorrentes

#=====
#caso2-Ajuste do MLG binomial com função de ligação logística com peso
#=====
w5 <- rep( 1/9.419 length(ida))
mod5<-glm(resp~rec+ida-1,weights=w5,family=binomial(link="logit")#rtas paralelas
w6 <- rep( 1/6.224, length(ida))
mod6<-glm(resp~rec/ida-1,weights=w6,family=binomial(link="logit")#retas concorrentes

#=====
#caso2-Ajuste do MLG binomial com função de ligação probito com peso w
#=====
w7 <- rep( 1/9.317, length(ida))
mod7<-glm(resp~rec+ida-1,weights=w7,family=binomial(link="probit")#retas paralelas
w8<- rep( 1/6.975, length(ida))
mod8<-glm(resp~rec/ida-1,weights=w8,family=binomial(link="probit")#retas concorrentes

#=====
#calculo de phi
#=====
phi1<- summary(mod1)$dispersion

#=====
#calculo de w
#=====
w9=1/phi1

#=====
#caso3-MLG binomial com função de ligação logística com peso e
eliminando (19 e 26)
#=====
w9 <- rep( 1/6.2685, length(ida))
mod9<-glm(resp~rec+ida-1,weights=w9,family=binomial(link="logit"),
```

```

subset=-c(19,26))#retas paralelas
w10 <- rep( 1/4.10, length(ida))
mod6<-glm(resp~rec/ida-1,weights=w10,family=binomial(link="logit"),
subset=-c(19,26))#retas concorrentes

#=====
#caso3-MLG binomial com função de ligação probito com peso w
#=====
w11 <- rep( 1/6.4381, length(ida))
mod11<-glm(resp~rec+ida-1, weights=w11,family=binomial(link="probit"),
subset=-c(19,26)) #retas paralelas
w12 <- rep( 1/4.8756, length(ida))
mod12<-glm(resp~rec/ida-1, weights=w12,family=binomial(link="probit"),
subset=-c(19,26))# retas concorrentes

#=====
#desvio residual e valor de p - caso 1 - ligação logística
#=====
(X2<-sum(residuals(mod1, 'deviance')^2))
1-pchisq(X2, df.residual(mod1))
(X2<-sum(residuals(mod2, 'deviance')^2))
1-pchisq(X2, df.residual(mod2))

#=====
#desvio residual e valor de p - caso 1 - ligação probito
#=====
(X2<-sum(residuals(mod3, 'deviance')^2))
1-pchisq(X2, df.residual(mod3))
(X2<-sum(residuals(mod4, 'deviance')^2))
1-pchisq(X2, df.residual(mod4))
#para os casos 2 e 3 é só substituir o objeto

#=====
#Análise de deviance- constroi-se a tabela usando os comandos
#=====
summary(mod1)
anova(mod1,test="Chisq")#se o teste for quiquadrado
anova(mod1,test="F")# caso o teste usado for F.

#=====
#os procedimentos apresentadas paro o caso 1,repetem-se para os casos 2 e 3.
#=====

#=====
#Gráfico - proporção de larvas x idade e curvas ajustadas.
#=====
(resp<-cbind(eclosao,total-eclosao))
pf<-eclosao/total
pfc<-split(prop,rec)
dc<-split(ida,rec)
plot(ida,prop,type="n",xlab="Idade dos ovos", ylab="Proporção de larvas")
points(dc[[1]],pfc[[1]],pch="*",col="blue")
points(dc[[2]],pfc[[2]],pch="+",col="green")
points(dc[[3]],pfc[[3]],pch="*",col="red")
lida<-seq(12,186,1)
lco<- predict(mod1,data.frame(ida=lida,rec=factor(rep("copo",length(lida)),
levels=levels(rec))),type="response")
lsa<- predict(mod1,data.frame(ida=lida,rec=factor(rep("enve",length(lida)),
levels=levels(rec))),type="response")
lev<- predict(mod1,data.frame(ida=lida,rec=factor(rep("saco",length(lida)),
levels=levels(rec))),type="response")
lines(lida,lco,lty=1,col="blue")
lines(lida,lsa,lty=1,col="green")
lines(lida,lev,lty=1,col="red")
legend(150,1,c("cop","saco","enve"),col=c("blue","green","red"),
lty=c(1,1,1),bty="o")

#=====
#Gráfico: proporção de larvas vesus idade e curvas ajustadas.
#=====
(resp<-cbind(eclosao,total-eclosao))
pf<-eclosao/total
pfc<-split(prop,rec)

```

```

dc<-split(ida,rec)
plot(ida,prop,type="n",xlab="Idade dos ovos", ylab="Proporção de larvas")
points(dc[[1]],pfc[[1]],pch="*",col="blue")
points(dc[[2]],pfc[[2]],pch="+",col="green")
points(dc[[3]],pfc[[3]],pch="*",col="red")
lida<-seq(12,186,1)
lco<- predict(mod2,data.frame(ida=lida,rec=factor(rep("copo",length(lida)),
levels=levels(rec))),type="response")
lsa<- predict(mod2,data.frame(ida=lida,rec=factor(rep("enve",length(lida)),
levels=levels(rec))),type="response")
lev<- predict(mod2,data.frame(ida=lida,rec=factor(rep("saco",length(lida)),
levels=levels(rec))),type="response")
lines(lida,lco,lty=1,col="blue")
lines(lida,lsa,lty=1,col="green")
lines(lida,lev,lty=1,col="red")
legend(150,1,c("cop","saco","enve"),col=c("blue","green","red"),
lty=c(1,1,1),bty="o")

#=====
#para os casos 2 e 3 é só substituir o modelo em questão.
#=====

#=====
#gráfico do ajuste versus resíduo desvio
#=====
plot(fitted(mod1), resid(mod1,type="deviance"), xlab="fitted",
ylab="Resíduos desvio",cex=0.7)

#=====
#gráfico qqnorm plot
#=====
qqnorm(resid(mod6,typr="deviance"),ylab="Resíduos desvio",
xlab="Quantis teóricos",cex=0.7,main="")
qqline(resid(mod6,type="deviance"))

#=====
#Teste para a função de ligação
#=====
fm1 <- (predict(mod1))^2)
mod12 <- update(mod1 , .~. +fm1, family=binomial)
mod12a <- update(mod1 , .~. +fm1, family=quasibinomial)
anova(mod12, test="Chisq")
anova(mod12a, test="F")

#=====
#gráficos para detectar pontos influêntes
#=====
influenceIndexPlot(mod1,vars=c("Cook","hat"))

#=====
# Para os outros envelopes é so mudar o nome do modelo.
#=====
#=====gráfico - envelope simulado=====
envel.bino <- function(modelo=mod1,iden=2,nome=seq(along =
model.matrix(modelo)[,1]),sim=100,conf=.90,res="D",quasi=F,
quad=T,maxit=20)
{
if(class(modelo)[1] != "glm") {
stop(paste("\nA classe do objeto deveria ser glm e nao ",class(modelo)
,"!!!\n"))
}
if(modelo$family[[1]] != "Binomial" & modelo$family[[1]] != "binomial" &
modelo$family[[1]]
!= "Aranda") {
if(modelo$family[[1]] == "quasibinomial" || ( modelo$family[[1]] == "Quasi-
likelihood" &
modelo$family[[3]] == "Binomial: mu(1-mu)" ) ) {
quasi<-T
} else {
stop(paste("\nA familia do objeto deveria ser Binomial e nao ",modelo$family
[[1]],"!!!\n"))
}
}
}

```

```

if(length(iden)>1) {
  iden<--1
}
alfa<-(1-conf)/2
X <- model.matrix(modelo)
n <- nrow(X)
p <- ncol(X)
if(is.null(modelo$prior.weights)) {
  ntot<-rep(1,n)
} else {
  ntot<-modelo$prior.weights
}
w <- modelo$weights
W <- diag(w)
H <- solve(t(X)%*%W%*%X)
H <- sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
h <- diag(H)
#para evitar divisão por 0 ao studentizar os resíduos, mas tentando
  manter o valor exagerado da alavanca h[round(h,15)==1]<-0.999999999999999
m<-predict(modelo,type="response")
y<-round(modelo$y*ntot,0)
fi <- 1
if(quasi==T) {
  fi <- 1/summary(modelo)$dispersion
}
if(res=="Q") {
  cat("Ao utilizar o resíduo Quantil para distribuições discretas,
  sugere-se plotar pelo menos 4 gráficos para evitar conclusões
  viesadas pela aleatoriedade que está sendo incluída.\n")
  tipo<-"Resíduo Quantil"
  r<-qnorm( runif(n=n,min=pbinom(y-1,ntot,m),max=pbinom(y,ntot,m))
    *sqrt(fi) )
} else {
  if(res=="D") {
    tipo<-"Resíduo Componente do Desvio"
    r<-resid(modelo,type="deviance")*sqrt(fi*(1-h))
  } else {
    if(res=="P") {
      tipo<-"Resíduo de Pearson Padronizado"
      r<-resid(modelo,type="pearson")*sqrt(fi*(1-h))
    } else {
      if(res=="W") {
        tipo<-"Resíduo de Williams"
        r<-sign((y/ntot)-m)*sqrt((1-h)*(( resid(modelo,type="deviance")*sqrt(fi/(1-h)) )
          ^2)+(h*( resid(modelo,type="pearson")*sqrt(fi/(1-h)) )^2))
      } else {
        if(res=="C") {
          tipo<-"Distância de Cook" r<-(h/((1-h)*p))*((resid(modelo,type="pearson")
            /sqrt(1-h))^2)
        } else {
          stop(paste("\nVoce nao escolheu corretamente um
          dos residuos disponiveis!!!\n"))
        }
      }
    }
  }
  link<-modelo$family[[2]]
  e <- matrix(0,n,sim)
  e1 <- numeric(n)
  e2 <- numeric(n)
  if (is.null(version$language) == T) {
    #No S-Plus, a opção start é para entrar com o preditor linear
    pm<-predict(modelo)
  } else {
    #No R, a opção start é para entrar com os coeficientes
    pm<-coef(modelo)
  }
  mu<-m
  for(i in 1:sim) {
    resp <- rbinom(n,ntot,mu)
    if ( (is.null(version$language) == T && link == "Logit:
    log(mu/(1 - mu))" ) | (is.null(version$language) == F && link ==

```

```

"logit") )
{
fit <- glm(cbind(resp,ntot-resp) ~ X-1,family=binomial,maxit=maxit,
start=pm)
} else {
if ( (is.null(version$language) == T && link == "Probit:
qnorm(mu)" ) | (is.null(version$language) == F && link == "probit") ) {
fit <- glm(cbind(resp,ntot-resp) ~ X-1,family=binomial(link=probit),maxit
=maxit,start=pm)
} else {
if ( (is.null(version$language) == T && link == "Complementary Log:
log( - log(1 - mu))" ) |
(is.null(version$language) == F && link == "cloglog") ) {
fit <- glm(cbind(resp,ntot-resp) ~ X-1,family=binomial(link=cloglog),
maxit=maxit,start=pm)
} else {
if ( modelo$family[[1]] == "Aranda" ) {
fit <- do.call("glm",list(cbind(resp,ntot-resp)~
X-1,family=aranda(as.numeric(link)),maxit=maxit,start=pm))
} else {
stop(paste("\nEsta funcao so aceita as ligacoes:
logito, probito, complementar log-log e Aranda-Ordaz !!!\nLigacao ",link,"
desconhecida!!!\n"))
}
}
}
}
w <- fit$weights
W <- diag(w)
H <- solve(t(X)%*%W%*%X)
H <- sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
h <- diag(H)
h[round(h,15)==1]<-0.9999999999999999
m <- predict(fit,type="response")
y <- round(fit$y*ntot,0)
e[,i] <-
sort( if(res=="Q") {
qnorm( runif(n=n,min=pbinom(y-1,ntot,m),max=pbinom(y,ntot,m)) )
} else {
if(res=="D") {
resid(fit,type="deviance")/sqrt(1-h)
} else {
if(res=="P") {
resid(fit,type="pearson")/sqrt(1-h)
} else {
if(res=="W") {
sign((y/ntot)-m)*sqrt((1-h)*(( resid(fit,type="deviance")/sqrt(1-h) )^2)+
(h*( resid(fit,type="pearson")/sqrt(1-h) )^2))
} else {
if(res=="C") { (h/((1-h)*p))*((resid(fit,type="pearson")/sqrt(1-h))^2)
} else {
stop(paste("\nVoce nao escolheu corretamente um dos residuos disponiveis!!!\n"))
}
}
}
}
}
}
for(i in 1:n) {
eo <- sort(e[i,])
e1[i] <- quantile(eo,alfa)
e2[i] <- quantile(eo,1-alfa)
}
med <- apply(e,1,median)
if(quad==T) {
par(pty="s")
}
if(res=="C") {
#Segundo McCullagh e Nelder (1989, pág.407) e Paula (2003, pág.57) deve-se usar
qnorm((n+1:n+.5)/(2*n+1.125))
#Segundo Neter et alli (1996, pág.597) deve-se usar qnorm((n+1:n-.125)/(2*n+0.5))
qq<-qnorm((n+1:n+.5)/(2*n+1.125))
plot(qq,sort(r),xlab="Quantil Meio-Normal",ylab=tipo, ylim=range(r,e1,e2), pch=16)

```

```

} else {
qq<-qnorm((1:n-.375)/(n+.25))plot(qq,sort(r),xlab="Quantil da Normal Padrão",
ylab=tipo, ylim=range(r,e1,e2), pch=16)
}
lines(qq,e1,lty=1,col="red")
lines(qq,e2,lty=1,col="red")
lines(qq,med,lty=2)
nome<-nome[order(r)]
r<-sort(r)
while((!is.numeric(iden))||(round(iden,0)!= iden)|| (iden < 0))
{
  cat("Digite o num.de pontos a ser identificado (0=nenhum) e <enter> para continuar\n")
out <- readline()
iden<-as.numeric(out)
}
if(iden>0) {identify(qq,r,n=iden,labels=nome)}
if(quad=="T") {
par(pty="m")
}
cat("Banda de ",conf*100,"% de confianca, obtida por ",sim," simulacoes.\n")
}
#####
#Chama o gráfico de probabilidade normal padrão com envelope simulado
#####
envel.bino(mod1,sim=100,conf=.90,maxit=50,res="D")

```