



UNIVERSIDADE ESTADUAL DA PARAÍBA
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

Rafaela Silveira Rodrigues Almeida

**Comparação de mutações no DNA com o uso
de uma biomembrana utilizando o modelo linear
generalizado**

Campina Grande - PB

Abril de 2016

Rafaela Silveira Rodrigues Almeida

Comparação de mutações no DNA com o uso de uma biomembrana utilizando o modelo linear generalizado

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Estatística Aplicada do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de especialista em Estatística.

Orientador: Prof. Dr. Tiago Almeida de Oliveira

Campina Grande - PB

Abril de 2016

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

A447c Almeida, Rafaela Silveira Rodrigues.
Comparação de mutações no dna com o uso de uma biomembrana utilizando o modelo linear generalizado [manuscrito] / Rafaela Silveira Rodrigues Almeida. - 2016.
32 p. : il. color.

Digitado.

Monografia (Especialização em Estatística Aplicada) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2016.

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

1. Mutação celular. 2.Família exponencial. 3. Modelos lineares Generalizados. 4. Dados genéticos. I. Título.

21. ed. CDD 519.535 2

Rafaela Silveira Rodrigues Almeida

Comparação de mutações no DNA com o uso de uma biomembrana utilizando o modelo linear generalizado

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Estatística Aplicada do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de especialista em Estatística.

Trabalho aprovado em 20 de abril de 2016.

BANCA EXAMINADORA



Prof. Dr. Tiago Almeida de Oliveira
Universidade Estadual da Paraíba



Prof. Dr. Gustavo Henrique Esteves
Universidade Estadual da Paraíba



Prof. Dr. Ricardo Alves de Olinda
Universidade Estadual da Paraíba

*Ao meu Deus, pela perseverança e discernimento que me concedestes ao longo deste curso.
Muito obrigada, Senhor Jesus Cristo!*

Agradecimentos

Aos meus amados pais, Vicente e Tânia, pelo apoio e carinho!

Ao meu esposo Robson e minha filha Lorena, pelo companheirismo e amor nas horas difíceis!

A minha irmã, Daniela, pela amizade e confiança que temos uma para com a outra!

Ao meu orientador, Dr. Tiago Oliveira de Almeida, pelos muitos conhecimentos transmitidos, ajuda e orientação. Para mim, a estatística se tornou mais simples com os seus ensinamentos! Muito obrigada!

Aos professores da Pós-Graduação em Estatística Aplicada (UEPB). Todos vocês contribuíram grandemente para esta minha formação, e eu me sinto muito grata!

Aos meus colegas de curso, Adrícia, Aline, Allana, Erivan, Isaac, Luzidark, Maria da Conceição, Raquel e Tatiana, pela amizade e parceria que estabelecemos ao longo deste curso!

“Não sobrevive a espécie mais forte, mas a que se adapta à mudança. ”
(Charles Darwin)

Resumo

A mutação celular é a alteração de um ou mais pares de base do DNA. Este trabalho teve como objetivo utilizar o Modelo Linear Generalizado (MLG) para análises dos dados de testes de micronúcleo, relacionando o surgimento de mutações com o uso de uma membrana, como também propor modelos de MLGs a este banco de dados, afim de analisá-los e compará-los. Os dados foram coletados a partir de um Teste de Micronúcleo provenientes do Laboratório de Biogenética da Universidade Estadual da Paraíba (UEPB). Foram coletadas amostras de sangue de 15 camundongos adultos da espécie *Mus musculus* (Swiss albino), três grupos de tratamentos, cada um com 5 animais. Sendo um grupo controle negativo, um grupo controle positivo lote 1 e um grupo controle positivo lote 2 os dois últimos grupos utilizaram-se a biomembrana depois da cirurgia. Foram propostos três modelos de MLG, que diferiram quanto a distribuição da variável resposta, Modelo I (distribuição Poisson), Modelo II (distribuição quasi-Poisson), Modelo III (distribuição Binomial Negativa). Os resultados mostraram, nos três modelos propostos, que as variáveis tempo e tempo elevado ao quadrado foram significativas para o número de micronúcleos, e que as demais variáveis não foram significativas. Dentre os três modelos, apenas o Modelo III, foi bem ajustado, possuindo um p-valor no teste de χ^2 de $0,05 < 0,1048635$. Utilizando o teste da razão de verossimilhança no Modelo III, observou que parâmetro da variável tempo elevado ao quadrado é necessário ao modelo, no qual sua exclusão leva a rejeição da hipótese nula do teste da Verossimilhança, pois seu p-valor $0,05 > 0,005677256$. Na análise de resíduos do Modelo III, considera-se que este está bem ajustado e que no gráfico da distância de Cook não se encontrou presença de ponto discrepante, ou seja, nenhum ultrapassou 0,30.

Palavras-chaves: Mutações, família exponencial, modelos lineares generalizados

Abstract

The cell mutation is the alteration of one or more base pairs of DNA. This study aimed to use the Generalized Linear Model (GLM) for analysis of micronucleus test data, relating the emergence of mutations with the use of a membrane, as well as propose models MLGs to this database in order to analyzed them and compare them. Data were collected from a micronucleus test from the Biogenetic Laboratory of the State University of Paraíba (UEPB). Blood samples were collected from 15 adult mice of the species *Mus musculus* (Swiss albino), three treatment groups, each with 5 animals. As a negative control, a positive control group lot one and lot two positive control group the latter two groups were used in the biomembrane after surgery. three models of MLG were proposed, which differ for the distribution of the response variable, Model I (Poisson distribution) Model II (quasi-Poisson distribution), Model III (negative binomial distribution). The results showed, the three proposed models, the variables time and high time squared was significant for the number of micronuclei, and the other variables were not significant. Of the three models, only the Model III was well adjusted, having a p-value in the χ^2 test $0,05 < 0,1048635$. Using the likelihood ratio test in Model III, noted that the parameter squared time variable is needed for the model in which their exclusion leads to rejection of the null hypothesis test Likelihood because their p-value $0,05 > 0,005677256$. In the Model III residue analysis, it is considered that it is well adjusted and in Cook distance graph not met the presence of discrepant point, ie no exceeded 0,30.

Key-words: Generalized Linear Model; Logistic Regression; Stomach Cancer and Esophagus

Lista de ilustrações

Figura 1 – Curvas da distribuição Poisson com diferentes médias	23
Figura 2 – Histograma da frequência do número de Micronúcleos nos tratamentos “lt1” (lote 1); “lt2” (lote 2); “Test” (testemunha) e “all” (todos os tratamentos).	25
Figura 3 – Verificação se os dados do Teste de Micronúcleo seguem uma distribuição Poisson	26
Figura 4 – Verificação se os dados do Teste de Micronúcleo seguem uma distribuição Poisson	26
Figura 5 – Interação entre tempo e médias de Micronúcleos nos tratamentos “lt1” (lote 1); “lt2” (lote 2); “Test” (testemunha)	27
Figura 6 – Gráficos dos resíduos para o Modelo III	29

Lista de tabelas

Tabela 1 – Notação de algumas distribuições da família exponencial	14
Tabela 2 – Função de ligação para algumas distribuições do tipo exponencial . . .	16
Tabela 3 – Equações para calcular o desvio de membros da família exponencial . .	20
Tabela 4 – Estatística descritiva dos dados do Teste de Micronúcleo	25
Tabela 5 – Análise de Variância dos tratamentos utilizando o Teste de Micronúcleo do Modelo I	28
Tabela 6 – vAnálise de Variância dos tratamentos utilizando o Teste de Micronúcleo do Modelo II	28
Tabela 7 – vAnálise de Variância dos tratamentos utilizando o Teste de Micronúcleo do Modelo III	28
Tabela 8 – Estatística descritiva dos dados do Teste de Micronúcleo	29

Sumário

1	INTRODUÇÃO	11
2	FUNDAMENTAÇÃO TEÓRICA	13
2.1	Modelo lineares generalizados	13
2.2	Componentes do Modelo Linear Generalizado	14
2.3	Estimação dos parâmetros do MLG	16
2.4	Medidas de qualidade do ajuste	19
2.4.1	Teste da Razão da Verossimilhança	19
2.5	Estatística de Pearson generalizada	20
2.6	Análise dos resíduos	21
2.6.1	Tipos de resíduos	21
3	MATERIAL E MÉTODOS	22
4	RESULTADOS E DISCUSSÃO	25
5	CONCLUSÃO	30
6	REFERÊNCIAS	31

1 Introdução

A mutação é a alteração de um ou mais pares de base do DNA, que podem ocorrer tanto em nível gênico, como cromossômico, em células germinativas ou somáticas. As mutações ocorrem, algumas vezes, devido as células serem expostas a agentes mutagênicos (VOGELSTEIN e KINZLER, 2004)

Os agentes mutagênicos têm a capacidade de gerarem mutações no material genético, no qual estes danos não são reparados na replicação celular, e podem ser classificados em; agentes químicos- substâncias que alteram as ligações químicas, ou substituem nucleotídeos por moléculas similares ,ex; os radicais livres; agentes físicos- são capazes de danificar as ligações químicas entre os nucleotídeos, muitas vezes não só causando mutação, mas também a destruição da cadeia de DNA, resultando em morte celular, ex. radiação ionizante e o raios UVC e UVB e agentes biológicos- ação de vírus e bactérias, que inoculam parte do DNA da célula que estão hospedando ou ocasionando falhas genéticas. (BRITT, 1996; WOGAN et al, 2004; KAMBARA et al, 2005)

Avaliar corretamente a relação do potencial mutagênico de um determinado agente, algumas vezes não só depende dos métodos laboratoriais analíticos, mas também da análise estatística realizada (TAJIMA, 1989; ROZAS et al.,2003). Diversos estudos genéticos comprovam a importância do método estatístico escolhido para uma conclusão precisa, STEPHENS et al, (2001) utilizaram o método Bayesiano para inferir a fase e reconstruir haplótipos que houveram mutações; FREI e WÜRGLER (1988) determinaram a mutagenicidade de substâncias produzidas por *Drosophila* comparando dois testes estatísticos, o binomial condicional e o teste 2 para proporções; e GU (1999) que avaliou a distinção entre genes por causa de mutações de membros do mesmo cluster utilizando o Modelo Markov.

Avaliar corretamente a relação do potencial mutagênico de um determinado agente, algumas vezes não só depende dos métodos laboratoriais analíticos, mas também da análise estatística realizada (TAJIMA, 1989; ROZAS et al.,2003). Diversos estudos genéticos comprovam a importância do método estatístico escolhido para uma conclusão precisa, STEPHENS et al, (2001) utilizaram o método Bayesiano para inferir a fase e reconstruir haplótipos que houveram mutações; FREI e WÜRGLER (1988) determinaram a mutagenicidade de substâncias produzidas por *Drosophila* comparando dois testes estatísticos, o binomial condicional e o teste ξ para proporções; e GU (1999) que avaliou a distinção entre genes por causa de mutações de membros do mesmo cluster utilizando o Modelo Markov.

Neste contexto, os modelos lineares generalizados (MLG) podem ser utilizados em

diversas abordagens genéticas, pois eles além de serem uma síntese de vários modelos estatísticos, podem estender as opções de distribuições de probabilidades dos erros associados ao modelo de regressão, flexibiliza a relação funcional, a variável resposta e as variáveis explicativas do modelo, o que muitas vezes são requeridos em dados genômicos. Assim, este estudo teve como objetivo utilizar o MLG para análises dos dados de testes de micronúcleo, relacionando o surgimento de mutações com o uso de uma membrana reticulada, como também ajustar modelos de MLGs a este banco de dados, afim de analisá-los e compará-los.

2 Fundamentação Teórica

O conteúdo desta seção relata os principais aspectos da utilização dos modelos lineares generalizados, com o objetivo de obter-se informações precisas sobre o assunto abordado.

2.1 Modelo lineares generalizados

Os modelos de regressão, introduzidos por Legendre e Gauss no início do século XIX, foram utilizados por muito tempo para descreverem a maioria dos fenômenos aleatórios, mesmo quando estes não apresentavam características de linearidade e alguns pressupostos requeridos deste modelo não eram alcançados, como por exemplo, a suposição de normalidade. Entretanto, algumas vezes, haviam formas de contornarem estes problemas utilizando algum tipo de transformação.

Posteriormente, vários modelos não lineares ou não normais foram desenvolvidos, tais como o modelo de regressão não-linear; o modelo complementar log-log para ensaios de diluição (FISHER, 1922), os modelos probit (BLISS, 1935) e logit (BERKSON, 1944; DYKE e PATTERSON, 1952; RASCH, 1960) para proporções, os modelos log-lineares para dados de contagens (BIRCH, 1963), os modelos de regressão para análise de sobrevivência (FEIGL e ZELEN, 1965; ZIPPIN e ARMITAGE, 1966; GLASSER, 1967).

Contudo, NELDER e WEDDERBURN (1972), propuseram um modelo, o Modelo Linear Generalizado (MLG), que é a síntese de vários modelos, com um processo iterativo para a estimação do parâmetro. Além disto, estes autores introduziram o conceito de desvio, que tem sido utilizado na avaliação de qualidade do ajuste do MLGs.

A elaboração de um MLG consiste na escolha de uma distribuição de probabilidade para a variável resposta que deve ser única e pertencente à família exponencial, das variáveis quantitativas e/ou qualitativas para representar a estrutura linear do modelo e de uma função de ligação. Adicionalmente, os termos que compõem a estrutura linear do modelo podem ser de natureza contínua, qualitativa ou mista, e devem contribuir significativamente na explicação da variável resposta. Deve-se haver suposição de independência, ou de não-correlação, entre as observações. Assim, dados exibindo autocorrelação no tempo, por exemplo, não devem fazer parte do contexto dos MLGs.

2.2 Componentes do Modelo Linear Generalizado

Os modelos lineares generalizados podem ser usados quando se tem uma variável aleatória Y e associado a ela um conjunto de variáveis explicativas, X_1, X_2, \dots, X_n . Para uma amostra de $n, (y_i, x_i)$ observações em que $(x_{1i}, x_{2i}, \dots, x_{pi})^T$, T é o vetor coluna de variáveis explicativas. Neste contexto, este modelo envolve três componentes:

- i) Componente aleatório – identificado pela variável aleatória resposta Y , em que esta variável é pertencente a distribuição da família exponencial. A função de densidade ou de probabilidade da família exponencial é dada por

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

em que, y é a variável resposta e $a(\phi)$, $b(\theta)$ e $c(y, \phi)$ são funções denominadas de funções específicas. O parâmetro θ é o parâmetro de localização e ϕ é o parâmetro de dispersão, associado à variância. Para se obter a média e a variância da variável y , deve-se calcular as derivadas de primeira e segunda ordem da função $b(\theta)$, mostrados a seguir

$$E(y) = \mu = \frac{db(\theta)}{d\theta},$$

e

$$var(y) = \frac{d^2b(\theta)}{d\theta^2} a(\phi). \tag{2.2}$$

Assim, a variância de y é um produto que depende de sua média e de $a(\phi)$. A parcela correspondente à segunda derivada de $b(\theta)$ é conhecida como função da variância, $V(\mu)$, que é dada pela equação 2.3

$$V(\mu) = \frac{var(y)}{a(\phi)} = \frac{d^2b(\theta)}{d\theta^2} = \frac{d\mu}{d\theta} \tag{2.3}$$

As principais distribuições pertencentes à família exponencial são apresentadas na Tabela 1

Tabela 1 – Notação de algumas distribuições da família exponencial

Distribuição	$a(\phi)$	θ	$b(\theta)$	$c(y; \phi)$	$\mu(\theta)$	$V(\mu)$
Normal	σ^2	μ	$\frac{\theta^2}{2}$	$-\frac{1}{2} \left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2) \right]$	θ	1
Poisson	1	$\ln \mu$	e^θ	$-\ln y!$	θ	μ
Binomial Negativa	1	$\ln \left(\frac{\mu}{1-k} \right)$	$-k \ln(1 - e^\theta)$	$\ln \left[\frac{\Gamma(k+y)}{\Gamma(k)y!} \right]$	$k \frac{e^\theta}{1-e^\theta}$	$\mu \left(\frac{\mu}{k} + 1 \right)$
Binomial	1	$\ln \left(\frac{\pi}{1-\pi} \right)$	$-m \ln(1 + e^\theta)$	$\ln \binom{m}{y}$	$m \frac{e^\theta}{1+e^\theta}$	$\frac{1}{m} \mu (m - \mu)$

- ii) Componente estrutural ou sistemático especifica as variáveis explicativas do modelo linear generalizado, ou covariáveis do modelo. As covariáveis (X_1, \dots, X_p) produzem uma estrutura linear η com caráter preditivo dada por

$$\eta = \beta_0 + \sum_{j=1}^p X_j \beta_j,$$

em que, $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ e o vetor que consiste dos coeficientes de regressão, frequentemente desconhecidos. Portanto, a componente sistemática do modelo pode ser escrita na forma:

$$\eta = Z_i^T \beta.$$

Sendo o β o vector dos coeficientes de regressão, em que:

$$z_i = (1, x_i^T)^T.$$

Na notação matricial,

$$\eta = Z\beta,$$

em que, Z é a matriz de especificação de dimensão $n \times (p + 1)$, e β é o vetor dos parâmetros de regressão de dimensão $(p + 1)$.

- iii) Função de ligação estabelece a ligação entre as componentes aleatória e sistemático, sendo uma função g , monótona e diferenciável, que relaciona o preditor linear h ao valor esperado μ da variável resposta

$$\eta = g(\mu).$$

Entretanto, deve-se levar em consideração o tipo de função de ligação escolhida, para que haja compatibilidade com a estrutura do erro e a interpretação do modelo. No modelo de regressão linear clássico, a ligação utilizada é a identidade, onde os valores esperados dos dados e os preditores lineares podem ter qualquer real. Nos MLGs, as funções de ligações canônicas são usualmente utilizadas. Nessas funções, o preditor linear é igual ao parâmetro canônico θ e têm-se uma ligação canônica sob a forma de identidade do tipo:

$$\theta_i = n_i,$$

em que

$$L(\beta) = g^{-1}(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}).$$

A escolha de uma ligação canônica simplifica as estimativas de máxima verossimilhança dos parâmetros do modelo e o cálculo do intervalo de confiança para a média da resposta. As ligações canônicas usualmente mais utilizadas nos MLGS estão expostas na Tabela 2.

Tabela 2 – Função de ligação para algumas distribuições do tipo exponencial

Distribuição	Ligação h
Poisson $P(\lambda)$	$h(\lambda) = \log(\lambda)$
Bernoulli $B(\pi)$	$h(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$
Binomial $B(M, \pi)$	$h(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$
Normal $N(\mu, \sigma)$	$h(\mu) = \mu$

2.3 Estimação dos parâmetros do MLG

O ajuste de um modelo linear generalizado é determinado pelo vetor de estimativas dos parâmetros. Existem diversos métodos que podem ser utilizados para a obtenção das estimativas dos β s, como o qui-quadrado mínimo, o Bayesiano e o método da máxima verossimilhança. O método mais utilizado é o método da máxima verossimilhança, o qual, o logaritmo da função de verossimilhança para um conjunto de observações independentes y_1, y_2, \dots, y_n é dado pela soma das contribuições individuais, isto é:

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

sendo que,

$$E(Y_i) = \mu_i, \eta_i = g(\mu_i) = x_i^T \beta \quad e \quad \theta_i = \int_f^V^{-1} d\mu_i = q(\mu_i).$$

Uma propriedade da família exponencial de distribuições é que seus elementos satisfazem a condições de regularidade suficientes para assegurar que o máximo global do logaritmo da função de verossimilhança $\ell(\boldsymbol{\theta}; \mathbf{y})$ é dado unicamente pela solução do sistema de equações $\mathbf{U}_\theta = \frac{d\ell}{d\theta} = \mathbf{0}$ equivalente $\mathbf{U}_\beta = \frac{d\ell}{d\beta} = \mathbf{0}$

Então a função score é

$$U_j = \sum_{i=1}^n \frac{\partial \ell(\theta_i; y_i, \phi)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \ell_i}{\partial \beta_j}.$$

No entanto,

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

$$\theta_i = \int_i^V^{-1} d\mu_i = q(\mu_i)$$

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

$$\eta_i = \sum_{j=1}^p x_{ij} \beta_j.$$

Tem-se, através da regra em cadeia

$$U_j = \sum_{i=1}^n \frac{d\ell_i}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^n \frac{d\ell_i}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j}.$$

Contudo,

$$E(Y_i) = \mu_i = b'(\theta_i) \text{ e } \text{Var}(Y_i) = a_i(\phi) \frac{d\mu_i}{d\theta_i} = a_i(\phi) V(\mu_i) \Rightarrow \frac{d\mu_i}{d\theta_i} = V(\mu_i).$$

Logo,

$$U_J = \frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \frac{1}{a_i(\phi)} (\gamma_i - \mu_i) \frac{1}{V(\mu_i)} \frac{d\mu_i}{d\eta_i} X_{ij} \quad (2.4)$$

Frequentemente, as equações $U_j = 0$, $j = 1, 2, \dots, p$ não são lineares e têm que ser resolvidas numericamente por processos iterativos do tipo Newton-Raphson. O método iterativo de Newton-Raphson para a solução de uma equação $f(x) = 0$ é baseado na aproximação de Taylor para a função $f(x)$ nas vizinhanças do ponto x_0 , ou seja,

$$f(x) \approx f(x_0) + (x - x_0) f'(x_0) = 0.$$

Obtendo-se,

$$x = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

ou, de forma mais geral

$$x^{(m+1)} = x^{(m)} - \frac{f(x^{(m)})}{f'(x^{(m)})}$$

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

Considerando-se que se deseja obter a solução do sistema de equações $\frac{d\ell}{d\beta} = 0$ e usando-se a versão multivariada do método de Newton-Raphson, tem-se

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \left(\mathbf{I}_0^{-1}\right)^{(m)} \mathbf{U}^{(m)}.$$

Sendo $\boldsymbol{\beta}^{(m)}$ e $\boldsymbol{\beta}^{(m+1)}$ os vetores de parâmetros estimados nos passos m e $(m+1)$, $\mathbf{U}^{(m)}$ o vetor escore, isto é, o vetor de derivadas parciais de 1ª ordem de $f(x)$, com elementos $\frac{\partial \ell}{\partial \beta_j}$, avaliado no passo m e $\left(\mathbf{I}_0^{-1}\right)^{(m)}$ a inversa da negativa da matriz de derivadas parciais de 2ª ordem de $f(x)$, com elementos $\frac{-\partial^2 \ell}{\partial \beta_j \partial \beta_k}$ avaliados no passo m .

Quando as derivadas de 2ª ordem são obtidas facilmente, o método de Newton-Raphson é bastante útil. Acontece, porém, que isso nem sempre ocorre e no caso dos modelos lineares generalizados usa-se o método escore de Fisher que, em geral, é mais simples (coincidindo com o método de Newton-Raphson no caso das funções de ligação canônicas). Ele envolve a substituição da matriz de derivadas parciais de 2ª ordem pela

matriz de valores esperados das derivadas parciais, isto é, a substituição da matriz de informação observada, pela matriz de informação esperada de Fisher, \mathfrak{S} . Logo,

$$\beta^{(m+1)} = \beta^{(m)} + (\mathfrak{S}^{-1})^{(m)} \mathbf{U}^{(m)} \quad (2.5)$$

sendo que \mathfrak{S} tem elementos dados por

$$\mathfrak{S}_{jk} = \mathfrak{S}^{(m)} \beta^{(m+1)} = \mathfrak{S}^{(m)} \beta^{(m)} + \mathbf{U}^{(m)}$$

que é a matriz de covariâncias dos \mathbf{U}_j s.

Multiplicando-se ambos os lados de (2.5) por tem-se

$$\mathfrak{S}^{(m)} \beta^{(m+1)} = \mathfrak{S}^{(m)} \beta^{(m)} + \mathbf{U}^{(m)} \quad (2.6)$$

Assim, obtém-se usado (2.6)

$$\begin{aligned} \mathfrak{S}_{jk} &= E(U_J U_K) = \sum_{i=1}^n \frac{1}{[a_i(\phi)]^2} E(Y_i - \mu_i)^2 \frac{1}{[V(\mu_i)]^2} \left[\frac{d\mu_i}{d\eta_i} \right]^2 x_{ij} x_{ik} \\ &= \sum_{i=1}^n \frac{1}{[a_i(\phi)]^2} a_i(\phi) V(\mu_i) \frac{1}{[V(\mu_i)]^2} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij} x_{ik} \\ &= \sum_{i=1}^n \frac{1}{[a_i(\phi)] [V(\mu_i)]} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij} x_{ik} \end{aligned}$$

e fazendo-se $a_i(\phi) = \frac{\phi}{w_i}$ com $\phi > 0$ constante, w_i peso a priori e $w_i = \frac{w_i}{V(\mu_i)} \left(\frac{d\mu_i}{d\eta_i} \right)^2$ tem-se

$$\mathfrak{S} = \frac{1}{\phi} \mathbf{X}^T \mathbf{W} \mathbf{X}$$

Com elementos,

$$\mathfrak{S}_{JK} = \sum_{i=1}^n \frac{1}{\phi} x_{ij} W_i x_{ik}$$

em que \mathbf{X} , a matriz do modelo e $\mathbf{W} = \text{diag}(W_1, W_2, \dots, W_n)$, No caso das funções de ligação canônicas $W_i = w_I V(\mu_1)$.

Além disso, rearranjando-se os termos de U_J tem-se

$$U_J = \sum_{i=1}^n \frac{w_i}{\phi} \frac{(y_i - \mu_i)}{V(\mu_i)} \frac{d\mu_i}{d\eta_i} x_{ij} = \sum_{i=1}^n \frac{w_i}{\phi} x_{ij} W_i \frac{d\eta_i}{d\mu_i} (y_i - \mu_i)$$

e, portanto, o vetor escore \mathbf{U} fica

$$\mathbf{U} = \frac{1}{\phi} \mathbf{X}^T \mathbf{W} \Delta (\mathbf{Y} - \boldsymbol{\mu})$$

com,

$$\Delta = \text{diag} \left\{ \frac{d\eta_1}{d\mu_1}, \frac{d\eta_2}{d\mu_2}, \dots, \frac{d\eta_n}{d\mu_n} \right\} = \text{diag} \left\{ g'(\mu_1), g'(\mu_2), \dots, g'(\mu_n) \right\}$$

Logo, substituindo \mathfrak{S} e \mathbf{U} em, tem-se

$$\frac{1}{\phi} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \beta^{(m+1)} = \frac{1}{\phi} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \beta^{(m)} + \frac{1}{\phi} \mathbf{X}^T \mathbf{W}^{(m)} \Delta^{(m)} (\mathbf{y} - \boldsymbol{\mu})^m$$

ou ainda,

$$\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} = \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m)} - \mathbf{Delta}^{(m)} (y - \mu)^m$$

e fazendo-se

$$\mathbf{Z}^{(m)} = \mathbf{X} \boldsymbol{\beta}^{(m)} + \mathbf{\Delta}^{(m)} (y - \mu)^m = \boldsymbol{\eta}^{(m)} + \mathbf{\Delta}^{(m)} (y - \mu)^m$$

chamada a variável dependente ajustada tem-se

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

ou ainda

$$\boldsymbol{\beta}^{(m+1)} = \left(\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^m \quad (2.7)$$

que tem a forma da solução das equações normais, para o modelo linear obtida pelo método dos quadrados mínimos ponderados, exceto que nesse caso a solução $\boldsymbol{\beta} = \boldsymbol{\beta}^{(m+1)}$ é obtida por processo numérico iterativo. O método usual para iniciar o processo iterativo é especificar uma estimativa inicial $\boldsymbol{\beta}^0$ e sucessivamente alterá-la até que a convergência seja obtida e, portanto, $\boldsymbol{\beta}$. Note, contudo que cada observação pode ser considerada como uma estimativa do seu valor médio, isto é, e, portanto, $\boldsymbol{\beta} = \boldsymbol{\beta}^{(m+1)}$. Adicionalmente, cada observação pode ser considerada como uma estimativa do seu valor médio, isto é $\hat{\mu}_i = y_i$, e, portanto,

$$\hat{\eta}_i = g(\hat{\mu}_i) = g(y_i)$$

2.4 Medidas de qualidade do ajuste

A etapa de inferência tem o objetivo de verificar a adequação do modelo e analisar as suas discrepâncias. Quando se ajusta um modelo a um conjunto de dados, pode se considerar, que se deseja substituir um conjunto de valores observados por um conjunto de valores ajustados $\hat{\mu}$ deduzidos do modelo (MCCULLAGH e NELDER,1989). Entretanto, os \hat{y}' s não são iguais aos y' s, existindo uma discrepância neste valor. Nos MLGs, existem formas de medir esta distância, podendo ser pela razão da máxima verossimilhança, conhecida como desvio, ou pela estatística de Pearson generalizada.

2.4.1 Teste da Razão da Verossimilhança

O desvio de um determinado modelo é definido como sendo o desvio deste modelo em relação ao modelo saturado

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

onde, L_{Mod} é a função de máxima verossimilhança do modelo em questão e L_{Sat} é a função de máxima verossimilhança do modelo saturado. O modelo saturado é o que tem n parâmetros,

um para cada observação. Ele atribui toda a variação ao componente sistemático e, portanto, ajusta-se perfeitamente, reproduzindo os próprios dados. μ , diferentemente do modelo nulo, que tem um único parâmetro, representado por um valor comum a todos os dados, atribuindo toda a variação entre os s ao componente aleatório. Pode-se então escrever:

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

Um modelo mal ajustado aos dados tem um grande desvio e um modelo bem adequado aos dados um baixo desvio. No entanto, um baixo desvio é difícil de se obter sem aumentar os números de parâmetros do modelo, assim o ideal é um modelo com poucas variáveis independentes tendo uma interpretação simples e um desvio moderado. Os desvios para os membros da família exponencial estão na Tabela 3.

Tabela 3 – Equações para calcular o desvio de membros da família exponencial

Distribuição	Desvio
Normal	$\sum_{i=2}^{\pi} (y_i - \hat{\mu}_i)^2$
Binomial Negativa	$2 \sum_{i=2}^{\pi} \left\{ y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) + (y_i + k) \log \left(\frac{\hat{\mu}_i - k}{y_i + k} \right) \right\}$
Binomial	$2 \sum_{i=2}^{\pi} \left\{ y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) + (n - y_i) \log \left(\frac{n - y_i}{n_i + \hat{\mu}_i} \right) \right\}$
Poisson	$2 \sum_{i=2}^{\pi} \left\{ y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right\}$

De forma mais sintética, para comparar a qualidade do ajustamento de dois modelos aplica-se o teste de razão de verossimilhanças sob a hipótese nula, sendo, H_0 : Os dois modelos têm a mesma qualidade de ajuste

A estatística de teste é dada por:

$$G = -2\ell_{M_1}(\beta) - (-2\ell_{M_2}(\beta)) \\ = -2 \left(\frac{\ell_{M_1}(\beta)}{\ell_{M_2}(\beta)} \right)$$

em que, $\ell_{M_1}(\beta)$ é a função log-verossimilhança do modelo M_1 e $\ell_{M_2}(\beta)$ a função log-verossimilhança do modelo M_2

A estatística de teste segue uma distribuição Qui-quadrado com $(p_2 - p_1)$ graus de liberdade

$$G \sim \chi_{p_2 - p_1}^2$$

2.5 Estatística de Pearson generalizada

A Estatística de Pearson generalizada é outra medida de adequabilidade de modelos frequentemente utilizada nos MLGs.

$$\chi^2 = \sum_{i=1}^n \frac{w_i(y_i, \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

em que, $V(\hat{\mu}_i)$ é a função de variância estimada para a distribuição do modelo em estudo.

2.6 Análise dos resíduos

A análise de resíduos é utilizada para avaliar a qualidade do ajuste de um modelo em relação a escolha da distribuição, da função de ligação, do preditor linear e também para identificar observações mal ajustada, e que por isso são mal explicadas pelo modelo (MCCULLAGH e NELDER, 1989). No contexto dos MLG's, seu uso é relacionado para avaliação da escolha da distribuição proposta para a variável resposta, da função de ligação e da função de variância, onde muitas vezes podem ter sido mal escolhidas devido à falta de uma análise exploratória dos dados. A análise de resíduos dos MLGs é semelhante a dos modelos lineares clássicos, mas possuem algumas adaptações. Na verificação da pressuposição de linearidade para o modelo linear clássico usam-se os vetores \mathbf{y} e $\boldsymbol{\mu}$ enquanto que para o modelo linear generalizado devem ser usados z , a variável dependente ajustada, e, o preditor linear. A variância residual, s_2 é substituída por uma estimativa consistente do parâmetro ϕ e a matriz \mathbf{H} torna-se:

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

No qual é equivalente a substituir por \mathbf{X} por $\mathbf{X}^{1fr m - \epsilon} \mathbf{W}$. Note-se que, agora que \mathbf{H} depende das variáveis explanatórias, da função de ligação e da função de variância, tornando mais difícil a interpretação da medida de leverage. Pode ser mostrado que:

$$\mathbf{V}^{-1/2} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \cong \mathbf{H} \mathbf{V}^{-1/2} (\mathbf{Y} - \boldsymbol{\mu})$$

sendo, $\mathbf{V} = \text{diag} \mathbf{V}(\mu_i)$. Isso mostra que \mathbf{H} mede a influência em unidades estudentizadas de \mathbf{Y} sobre $\boldsymbol{\mu}$

2.6.1 Tipos de resíduos

Alguns tipos de resíduos mais usados para os modelos lineares generalizados são:

- i) Resíduo ordinário: é dado a partir da solução de mínimos quadrados da regressão linear ponderada de \tilde{Z} contra \mathbf{X} , definida por

$$r^* = \hat{w}^{1/2} [z - \hat{\eta}] = \mathbf{V}^{-1/2} (\mathbf{y} - \hat{\boldsymbol{\mu}})$$

- ii) Resíduos de Pearson generalizados

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\frac{\hat{\phi}}{w_i} \mathbf{V}(\hat{\mu}_i)}}$$

sendo ϕ uma estimativa consistente do parâmetro ϕ e w_i e um peso a priori (na maior parte dos casos igual a 1).

3 Material e métodos

O estudo foi realizado com dados de teste de micronúcleo provenientes do Laboratório de Biogenética da Universidade Estadual da Paraíba (UEPB). Foram coletadas amostras de sangue de 15 camundongos adultos da espécie *Mus musculus* (Swiss albino) após uma cirurgia na parte dorsal dos animais, onde foram divididos em três grupos de tratamentos, cada um com 5 animais. Sendo um grupo controle negativo (não utilizou a biomembrana curativa após a cirurgia), um grupo controle positivo lote 1(0217121) e um grupo controle positivo lote 2 (060411H), os dois últimos grupos utilizaram-se a biomembrana depois da cirurgia. Após isto, foram anotadas as frequências de micronúcleos, analisando 2000 eritrócitos policromáticos anucleados (PCEs), nos tempos 30h, 48h e 720h. Neste contexto, como os dados são de contagens, propôs-se três Modelos Lineares Generalizados ao presente banco de dados, afim de analisa-los e compará-los.

O Modelo I considera os dados da variável resposta como pertencentes á distribuição Poisson, possuindo as seguintes características:

- a) O componente aleatório, Y , é o número de micronúcleos (com distribuição Poisson)
- b) Os componentes sistemáticos é X_1 =controle negativo; X_2 =lote1; X_3 = lote2; X_4 =tempo; $X_5 = tempo^2$

A função de ligação utilizada foi: $h(\lambda) = \log(\lambda)$ e a função de densidade da distribuição Poisson é:

$$\begin{aligned} f(y; \mu) &= \frac{\mu^y \exp^{-\mu}}{y!} \\ &= \mu^y \exp^{-\mu} (y!)^{-1} \\ &= \exp \left\{ -\mu + \log(\mu^y) + \log(y!)^{-1} \right\} \\ &= \exp \left\{ -\mu + y \log(\mu) - \log(y!) \right\} \\ &= \exp \left\{ y \log(\mu) - \mu - \log(y!) \right\} \end{aligned}$$

No qual, $\theta = \log(\mu)$, $\phi = 1$, $b(\theta) = e^\theta = \mu$ e $c(y, \phi) = \log(y!)$. A função de variância $V(\mu) = \mu$.

$$E(Y) = \frac{DB(\theta)}{D\theta} = \mu; \quad Var(Y) = \frac{d^2b(\theta)}{d\theta^2} = \mu$$

Segundo Tadano (2007), a medida que aumenta, há um achatamento da curva da distribuição de Poisson, e um deslocamento à direita, aproximando-se á uma curva da distribuição normal, como mostra a Figura 3.

O Modelo II propôs que o Y é uma quase- Poisson, por poder haver uma sobredispersão dos dados. De acordo com Santos (2013), a sobredispersão é quando variância da

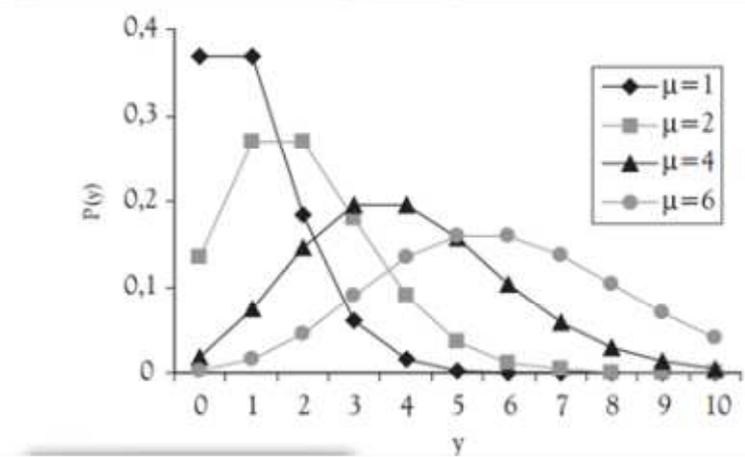


Figura 1 – Curvas da distribuição Poisson com diferentes médias

resposta é superior ao valor médio do fenômeno. Designando por ϕ o parâmetro de sobre-dispersão, tal que $Var(Y) = \phi IE(Y) = \phi$, quando ocorre sobre-dispersão na estimação dos parâmetros do modelo as estimativas pontuais, são iguais, caso não exista sobre-dispersão, mas a variância dos estimadores é inflacionada pelo parâmetro de sobre-dispersão ϕ

Assim, o Modelo II possui esta estrutura:

- O componente aleatório, Y , é o número de micronúcleos (com distribuição quase-Poisson)
- Os componentes sistemáticos é X_1 =controle negativo; X_2 =lote1; X_3 = lote2; X_4 =tempo; $X_5 = tempo^2$

A função de ligação: $h(\lambda) = \log(\lambda)$

O Modelo III considera que os dados da variável resposta, Y , é pertencente a distribuição Binomial Negativa.

Seja Y uma variável aleatória que segue uma distribuição Binomial Negativa com parâmetros k e p , $Y \sim BN(k, p)$. A variável Y representa o número de insucessos anteriores a k sucessos, num conjunto de acontecimentos independentes e com a mesma probabilidade de sucesso, p . A função densidade de probabilidade de Y é dada por:

$$\begin{aligned}
 f(y|k, p) &= \binom{y+k-1}{k-1} p^k (1-p)^y \\
 &= \exp \left[y \log(1-p) + k \log(p) + \log \binom{y+k-1}{k-1} \right]
 \end{aligned}$$

Neste caso, a distribuição Binomial Negativa está escrita na forma canônica, em

que:

$$\theta = \log(1 - p), b(\theta) = -k \log(p), a(\phi) = 1, ec(y, \phi) = \log \binom{y + k - 1}{k - 1}$$

A média e a variância são expressas, respectivamente, por:

$$E(Y) = b'(\theta) = \frac{k(1 - p)}{p}$$

e

$$Var(Y) = a(\phi) b(\theta) = \frac{k(1 - p)}{p^2}$$

O Modelo III apresenta a seguinte estrutura:

- a) O componente aleatório, Y , é o número de micronúcleos (com distribuição Binomial Negativa)
- b) Os componentes sistemáticos é X_1 =controle negativo; X_2 =lote1; X_3 = lote2; X_4 =tempo; $X_5 = tempo^2$

A função de ligação: $h(\pi) = \log(\pi/(1 - \pi))$

Todas as análises estatísticas foram feitas no Software R.Studio, e os pacotes utilizados foram o *foreign*, *ggplot2*, *MASS*, *vcd*, *boot*.

4 Resultados e discussão

A investigação do potencial mutagênico de compostos é de fundamental importância para evitar possíveis efeitos adversos para a saúde humana (ARENSEN et al., 2011). A estatística descritiva dos dados do Teste de Micronúcleo para avaliar a mutagenicidade da membrana se encontra na Tabela 4. O número médio de micronúcleos encontrados nas amostras foi de 1,689, entretanto, houveram amostras em que nenhum micronúcleo foi verificado, e algumas, 7 micronúcleos. Os histogramas de frequência do número de micronúcleos nos diferentes tratamentos estão na Figura 4.

Tabela 4 – Estatística descritiva dos dados do Teste de Micronúcleo

	Tempo	Amostra	Micronúcleo
Mínimo	30	319,0	0,000
Primeiro quartil	30	332,0	0,000
Mediana	48	343,0	1,000
Média	266	343,4	1,689
Terceiro quartil	720	356,0	2,000
Máximo	720	368,0	7,000

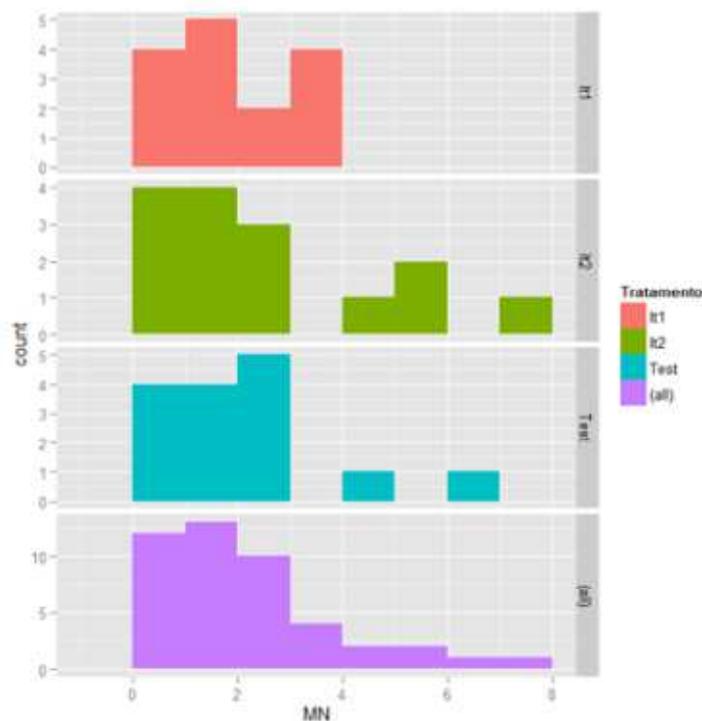


Figura 2 – Histograma da frequência do número de Micronúcleos nos tratamentos “It1” (lote 1); “It2” (lote 2); “Test” (testemunha) e “all” (todos os tratamentos).

Na Figura 4 observa-se na junção de todos os tratamentos, que os dados podem pertencer a uma distribuição Poisson ou Binomial Negativa, como propostos nos modelos deste estudo. No entanto, a avaliação de que as distribuições da variável resposta propostas nos modelos são de fato estas, foi feita através do Teste de χ^2 . Os resultados mostraram que os dados do número de micronúcleos podem pertencer a distribuição Poisson, pois encontrou-se p-valor $0,17 > 0,05$. Portanto, não rejeitando H_0 do Teste de χ^2 , que afirma que os dados são distribuídos segundo Poisson (Figura 4). Semelhantemente, o Teste de χ^2 para verificação se estes dados do Teste de Micronúcleo também podem seguir uma distribuição Binomial Negativa, não rejeitou H_0 , com p-valor de $0,78 > 0,05$. Assim, podendo também os dados serem caracterizados como uma Binomial Negativa (Figura 4).

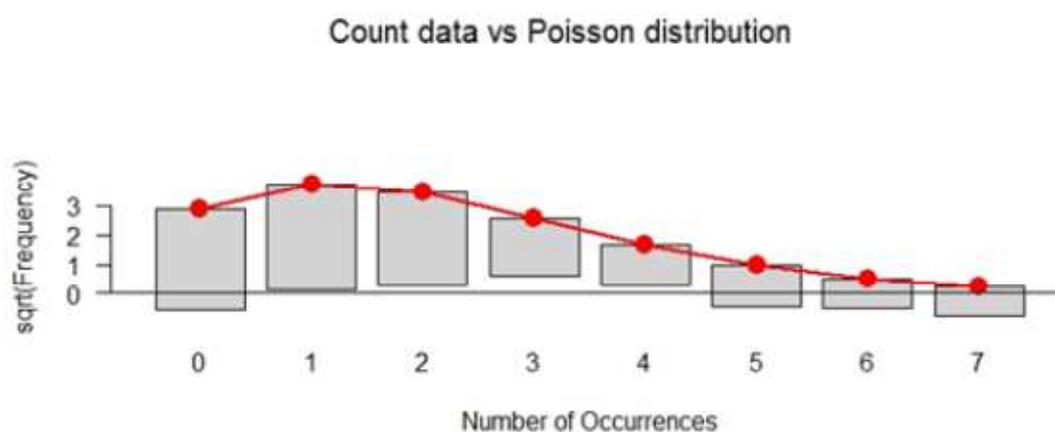


Figura 3 – Verificação se os dados do Teste de Micronúcleo seguem uma distribuição Poisson

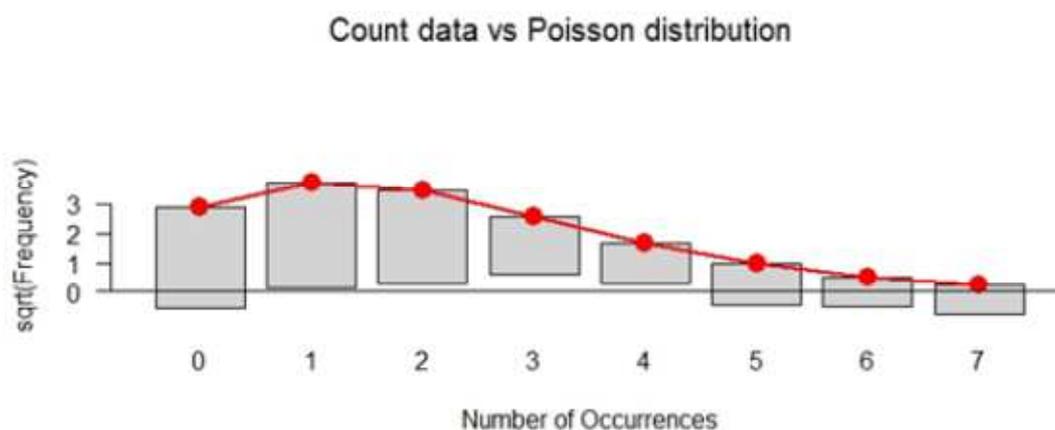


Figura 4 – Verificação se os dados do Teste de Micronúcleo seguem uma distribuição Poisson

A interação entre tempo e médias de micronúcleos nos tratamentos estão apresentadas na Figura 5. Analisando esta Figura, pode-se supor interações entre as variáveis

tempo nos tratamentos lote 1, lote 2 e testemunha. Entretanto, Segundo CORRAR e THEÓPHILO (2014), a análise gráfica muitas vezes não é precisa em algumas situações, necessitando de avaliações mais concretas para uma conclusão correta. Ainda avaliando a Figura 5, observa-se que o ápice do número de Micronúcleos se encontra em 48 horas na testemunha e após a implantação da membrana no tratamento lote 1, e diferentemente, no tratamento lote 2, o maior número de micronúcleos foi verificado após 720 horas da implantação.

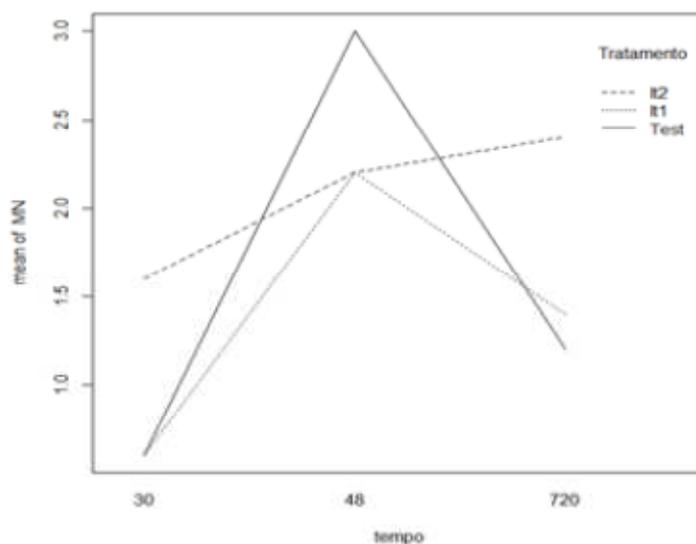


Figura 5 – Interação entre tempo e médias de Micronúcleos nos tratamentos “lt1” (lote 1); “lt2” (lote 2); “Test” (testemunha)

Em relação as estimativas das interações entre os tratamentos, no Modelo I (Tabela 5), em que a distribuição da variável resposta foi ajustada como pertencente a distribuição de Poisson, do Modelo II, em que Y pertence a uma quasi-Poisson (Tabela 6) e o Modelo III, que a variável resposta é pertencente a distribuição Binomial Negativa (Tabela 7), mostraram, nos três modelos, que as variáveis tempo e tempo elevado ao quadrado foram significativas para o número de micronúcleos, e que as demais variáveis não foram significativas.

Assim, a membrana avaliada não induz danos cromossômicos nos eritrócitos imaturos da espécie em estudo, pois não existem diferenças entre o número de micronúcleos do tratamento testemunha e os tratamentos que se utilizaram a membrana. Entretanto, o tempo foi uma variável que influenciou no número de micronúcleos nos tratamentos.

Em relação aos ajustes dos modelos, o deviance residual encontrado do Modelo I foi de 63,643 sobre 38 graus de liberdade, e o p-valor do teste de χ^2 foi de 0,005672327, ou seja, rejeita-se H_0 , pois o p - valor < 0,05. Assim, este modelo não está bem ajustado. O Modelo II, igualmente ao Modelo I, apresenta um deviance residual de 63,643 sobre 38 e um p-valor no teste de χ^2 de 0,005672327. Sendo assim, também, rejeitando H_0 . Contudo,

Tabela 5 – Análise de Variância dos tratamentos utilizando o Teste de Micronúcleo do Modelo I

Coeficientes	Estimativas	Erro padrão	Valor Z	Pr(> z)
(Intercepto)	-1,7053	0,8230	-2,07	0,0383
(Tratamento, "lt2")lt1	-0,2970	0,3757	-0,79	0,4292
(Tratamento, "lt2")Test	-0,0217	0,3512	-0,06	0,9508
Tempo	0,0603	0,0194	3,10	0,0019*
Tempo ²	-0,0001	0,0000	-3,11	0,0019*
Tempo:Tratamentolt2	0,0003	0,0009	0,37	0,7117
Tempo:TratamentoTest	-0,0006	0,0010	-0,60	0,5454

Tabela 6 – vAnálise de Variância dos tratamentos utilizando o Teste de Micronúcleo do Modelo II

Coeficientes	Estimativas	Erro padrão	Valor Z	Pr(> z)
(Intercepto)	-1,7053	1,0191	-1,67	0,1025
(Tratamento, "lt2")lt1	-0,2970	0,4652	-0,64	0,5270
(Tratamento, "lt2")Test	-0,0217	0,4348	-0,05	0,9605
Tempo	0,0603	0,0241	2,51	0,0166*
Tempo ²	-0,0001	0,0000	-2,51	0,0164*
Tempo:Tratamentolt2	0,0003	0,0011	0,30	0,7670
Tempo:TratamentoTest	-0,0006	0,0012	-0,49	0,6281

Tabela 7 – vAnálise de Variância dos tratamentos utilizando o Teste de Micronúcleo do Modelo III

Coeficientes	Estimativas	Erro padrão	Valor Z	Pr(> z)
(Intercepto)	-1,6910	0,9165	-1,85	0,0650
(Tratamento, "lt2")lt1	-0,3537	0,4387	-0,81	0,4201
(Tratamento, "lt2")Test	-0,0944	0,4185	-0,23	0,8215
Tempo	0,0610	0,0220	2,78	0,0055*
Tempo ²	-0,0001	0,0000	-2,78	0,0054*
Tempo:Tratamentolt2	0,0002	0,0010	0,24	0,8113
Tempo:TratamentoTest	-0,0006	0,0011	-0,51	0,6088

o Modelo III, que considera a variável Y pertencente a distribuição Binomial Negativa, foi bem ajustado, possuindo um p-valor no teste de χ^2 de $0,1048635 > 0,05$, aceitando H_0 . O deviance residual modelo foi de 49,227 sobre 38 de liberdade.

Neste contexto, pode-se concluir que o modelo que propunha a variável Y como pertencente a uma distribuição Binomial Negativa, foi o único que se ajustou aos dados. Segundo Paula (2007) em certas situações práticas, os modelos que consideram a binomial negativa como a distribuição de Y apresentam superioridade em relação aos que modelam com a distribuição Poisson, pois há melhor desempenho nos estimadores e controle da variabilidade dos dados, já que não necessitam que a variância seja igual a média.

Adicionalmente, para verificar se todos os parâmetros são relevantes para o ajuste do Modelo III, excluiu-se a variável tempo ao quadrado do modelo. Os resultados mostraram

que este parâmetro é necessário ao modelo, no qual sua exclusão leva a rejeição de H_0 do teste da Verossimilhança, pois seu valor p $0,05 > 0,005677256$, fazendo com que o Modelo III não fique mais ajustado, como mostrado na Tabela 8.

Tabela 8 – Estatística descritiva dos dados do Teste de Micronúcleo

Estimativa	G.L Residual	2 x log-lik	Test	G.L	LR stat	Pr(Chi)
2,848281	39	-154,13				
4,982437	38	-146,4823	1 vs	1	7,649992	0,005677256

Na Figura 4, estão os gráficos dos resíduos do Modelo III. Percebe-se que o modelo, também por esta análise, está bem ajustado, pela proximidade dos pontos ordenados a reta no gráfico do quantil do normal padrão e os resíduos desviante ordenados. No gráfico da distância de Cook, não há presença de ponto discrepante, ou seja, nenhum ultrapassou 0,30.

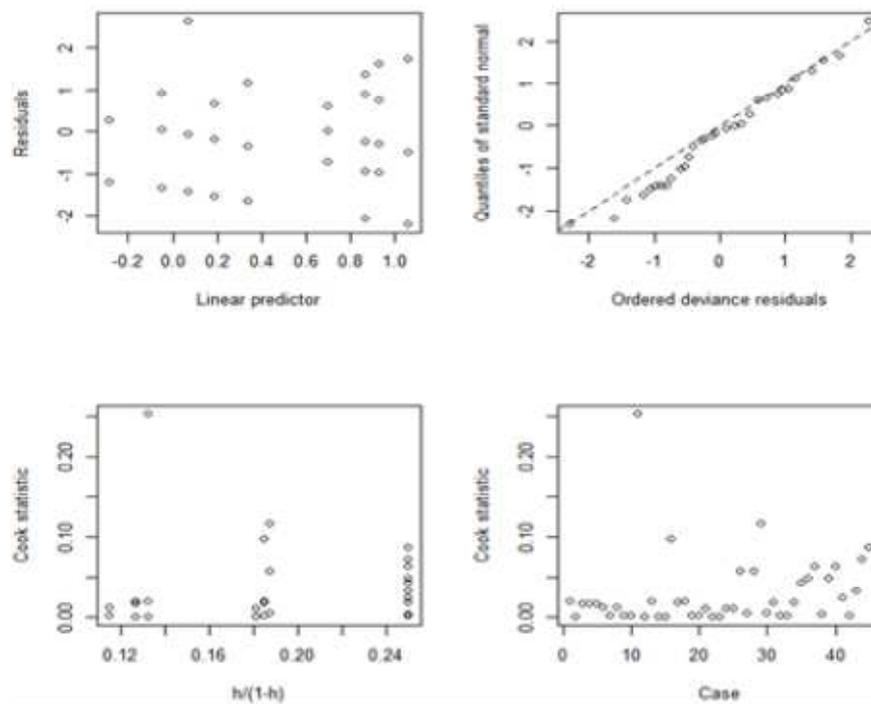


Figura 6 – Gráficos dos resíduos para o Modelo III

5 Conclusão

As estimativas dos modelos propostos mostraram que a variável tempo e tempo ao quadrado são significativas na mudança do número de micronúcleos encontrados nos eritrócitos da espécie *Mus musculus*. Contudo, não houveram diferenças significativas no número de micronúcleos da testemunha para os tratamentos que utilizaram a biomembrana. Assim, o uso da membrana estudada, possivelmente, não induz a mutagenicidade nas células.

O Modelo III, que considera a variável resposta como pertencente a distribuição Binomial Negativa, foi o único que se apresentou ajustado aos dados, contrariamente ao Modelo I, com distribuição Poisson no Y , e o Modelo II, que considerava a distribuição de Y como quase-Poisson, que não se ajustaram aos dados. Os resíduos do Modelo III se encontram bem ajustados, assim como no gráfico da distância de Cook, que não foram encontrados pontos discrepantes.

6 Referências

- ARENSON, A; PEREIRA NETO, T. J; GERBER, W. **Manual sobre toxicidade em efluentes industriais**. Porto Alegre: CEP SENAI de Artes Gráficas Henrique d'Ávila Bertaso, 2011.
- BERKSON, J. Application of the logistic function to bioassay. **J. R. Statist. Soc. B**, 39, 357-365, 1944.
- BIRCH, M. W. Maximum likelihood in three-way contingency tables. **J. R. Statist. Soc. B**, 25, 220-233, 1963.
- BLISS, C. I. The calculator of the dosage-mortality curve. **Ann. Appl. Biol.**, 22, 134-167, 1935.
- BRITT, A. B . DNA DAMAGE AND REPAIR IN PLANTS. **Annual Review of Plant Physiology and Plant Molecular Biology**, Vol. 47: 75-100 , 1996
- COSTA, V. C.; COSTA, H. M.; PEREIRA, M. M.; VASCONCELOS, W. L.; LAMBERT, R. O.; MANSUR, H. S. Preparation of Hybrid Biomaterials for Bone Tissue Engineering. **Materials Research** , v. v. 10, p. 21-26, 2007.
- CORRAR, L. J.; THEÓPHILO, C R. **Pesquisa Operacional para decisão em contabilidade e administração: Contabilometria**. 2ª edição. São Paulo, Atlas, 2014
- DYKE, G.; PATTERSON, H. Analysis of factorial arrangements when the data are proportions. **Biometrics**, 8, 1-12, 1952.
- FEIQL, P.; ZELEN, M. Estimation of exponential survival probabilities with concomitant information. **Biometrics**, 21, 826-838, 1965.
- FISHER, R. On the mathematical foundations of theoretical statistics. **Philosophical Transactions of the Royal Society**, 222, 309-368, 1922.
- GASSER, M. Exponential survival with covariance. **Journal of the American Statistical Association**, 62, 561-568, 1967
- GU, X. Statistical Methods for Testing Functional Divergence after Gene Duplication. **Mol. Biol. Evol.** 16(12):1664-1674. 1999.
- KAMBARA, H.; YOSHINAGA, S.EI.; CHIO, A.C. Cyclophosphamide Allows for In vivo Dose Reduction of a Potent Oncolytic Virus. **Cancer Res**, v. 15, 2005.
- FREI, H. ; WÜRGLER, F. E. Statistical methods to decide whether mutagenicity test data from *Drosophila* assays indicate a positive, negative, or inconclusive result. **Mutation Research**, 203: 297-308.1998.

- MCCULLAQH, P.; NELDER, J. A. **Generalized Linear Models**. Chapman and Hall, London, second edition. 1989
- NELDER, J. A.; WEDBERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society, A**, 135, 370-384, 1972.
- PAULA, G.A. **Modelos de Regressão com Apoio Computacional**, São Paulo: IME – Universidade de São Paulo. 2007.
- PEREIRA, A.P.V.; VASCONCELOS, W.L.; ORÉFICE, R.L. Novos Biomateriais: híbridos orgânico-inorgânicos bioativos. **Polímeros Ciência e Tecnologia**, n.4, p 104- 109. 1999.
- RASCH, G. **Probabilistic Models for Some Intelligence and Attainment Tests**. Danmarks Paedagogiske Institut, Copenhagen. 1960
- ROZAS, J.; BARRIO, J. C.S.; MESSEGUER, X.; ROZAS, R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. **Bioinformatics** 19(18), 2003
- STEPHENS, M.; SMITH, N.J.; DONNELLY, P. A New Statistical Method for Haplotype Reconstruction from Population Data. **The American Journal of Human Genetics**. v. 68, Issue 4, Pages 978–989, 2001
- TADANO, Y. I.S.; UGAYA, C.M.L.; FRANCO, A.T. Método de regressão de Poisson: metodologia para avaliação do impacto da poluição atmosférica na saúde populacional. **Ambiente & Sociedade**, Campinas v. XII, n. 2, p. 241-255. 2009
- TAJIMA, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics**, 123, 585–595, 1989.
- VOGELSTEIN, B.; KINZLER, K.W. Cancer genes and the pathways they control. **Nature Medicine**, 10, 789 - 799, 2004.
- ZIPPIN, C.; ARMITAGE, P. Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. **Biometrics**, 22, 665-672, 1966.
- WOGAN, G.N.; HECHT, S.S.; FELTONC, J.S.; CONNEY, A.H.; LOEB, L.A. **Environmental and chemical carcinogenesis**. Volume 14, Issue 6, Pages 473–486, 2004.