



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

Roseane de Alcântara Costa

Uso do modelo logístico multinomial para avaliar o efeito de adubação potássica no controle do tripses do prateamento

Campina Grande - PB

7 de Dezembro de 2017

Roseane de Alcântara Costa

**Uso do modelo logístico multinomial para avaliar o efeito
de adubação potássica no controle do tripses do
prateamento**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística 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 bacharel em Estatística.

Orientador: Profa. Dra. Ana Patricia Bastos Peixoto

Coorientador: Profa. Dra. Gláucia Amorim Faria

Campina Grande - PB

7 de Dezembro de 2017

É expressamente proibido 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 do trabalho.

C837u Costa, Roseane de Alcantara.

Uso do modelo logístico multinomial para avaliar o efeito de adubação potássica no controle do tripses do prateamento [manuscrito] : / Roseane de Alcantara Costa, Profa. Dra. Gláucia Amorim Faria. - 2017.

37 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2017.

"Orientação : Profa. Dra. Ana Patricia Bastos Peixoto, Departamento de Estatística - CCT."

1. Enneothrips flavens. 2. Modelo multinomial . 3. Cultura do amendoim. 4. Tripses do prateamento.

21. ed. CDD 519.5

Roseane de Alcântara Costa

Uso do modelo logístico multinomial para avaliar o efeito de adubação potássica no controle do trips do prateamento

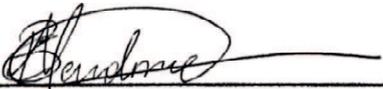
Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Trabalho aprovado em 7 de Dezembro de 2017.

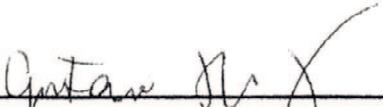
BANCA EXAMINADORA



Profa Dra. Ana Patricia Bastos Peixoto
Universidade Estadual da Paraíba



Prof. Me. Edúrio Barbosa de Mendonça
Universidade Federal de Campina Grande



Prof. Dr. Gustavo Henrique Esteves
Universidade Estadual da Paraíba

Agradecimentos

Agradeço em primeiro lugar a Deus por ter me dado forças e sabedoria, para a conclusão desta etapa da minha vida.

Ao corpo docente do Departamento de Estatística da UEPB, em especial a minha orientadora Doutora Ana Patricia, por ter me orientado na conclusão deste trabalho.

Ao meu colegas de turma, e aos demais companheiros com quem estive em outras disciplinas.

Ao professor Luciano Albino, por todos os ensinamentos e conselhos que somaram em minha vida acadêmica, pela sua amizade e parceria.

A minha tia Marlene, que sempre compreendeu a minha ausência e sempre torceu por mim.

As minhas amigas Valdirene, Mariana e Aucicleide que sempre respeitaram minhas escolhas por mais difíceis que elas fossem.

Ao meu esposo e amigo Manoel por estar sempre ao meu lado me apoiando em todas as escolhas da minha vida, e que sempre torce para o meu melhor.

E a todos que contribuíram de alguma forma para a conclusão deste trabalho.

*“O futuro dependerá daquilo que fizemos no presente”
(Mahatma Gandhi)*

Resumo

O tripses do prateamento, *Enneothrips flavens*, é considerado a principal praga do amendoim no Brasil. Por sua ocorrência generalizada, os danos causados pelos tripses varia de acordo com o estágio de crescimento e, conseqüentemente, o impacto econômico está associado com a idade fisiológica da planta, pelos elevados níveis populacionais e danos causados a cultura. Esse trabalho teve por objetivo utilizar modelos lineares generalizados para avaliar o comportamento das doses de potássio sobre o manejo do tripses e produtividade da cultura do amendoim. Avaliaram-se os tratamentos provenientes de um experimento, cujo delineamento foi em blocos casualizados no esquema de parcelas subdivididas no tempo, para avaliar como tratamentos as doses de 0, 50, 100, 150 e 200 *kg/ha* de potássio, aplicado em cobertura, 8 dias após emergência das plantas, com 5 repetições. O modelo ajustado aos dados foi o multinomial, apresentando-se adequado e os resultados afirmam que o potássio está envolvido no mecanismo de defesa das plantas a pragas e doenças, de modo que, altas concentrações de potássio promovem a síntese e o acúmulo de compostos fenólicos, inibidores de insetos e fungos.

Palavras-chaves: *Enneothrips flavens*; Modelos multinomial; Cultura do amendoim.

Abstract

The silver thrips, *Enneothrips flavens*, are considered the main plague of peanuts in Brazil, due to its widespread occurrence, the damage caused by the thrips varies according to the stage of growth and, consequently, the economic impact is associated with with the physiological age of the plant, by the high population levels and damages caused to culture. The objective of this work was to use generalized linear models to evaluate the behavior of potassium doses on thrip management and productivity of peanut culture. The treatments were evaluated from an experiment, whose design was a randomized complete block design subdivided in the time, to evaluate as treatments the doses of 0, 50, 100, 150 and 200 kg ha^{-1} of potassium, applied on cover, 8 days after emergence of the plants, with 5 replicates. The model adjusted to the data was the multinomial, presenting itself adequate and the results affirm that potassium is involved in the plant defense mechanism of pests and diseases, Thus, high concentrations of potassium promote the synthesis and accumulation of compounds phenolics, insect and fungi inhibitors.

Key-words: *Enneothrips flavens*; Multinomial models; Peanut Culture.

Lista de ilustrações

Figura 1 – Boxplot dos dados da cultura do Amendoim.	26
Figura 2 – Modelo ajustado pela distribuição Poisson.	27
Figura 3 – Gráfico de probabilidade ajustado para os dias de Avaliação.	29
Figura 4 – Gráfico de risco ajustado pela distribuição Multinomial.	30
Figura 5 – Modelo ajustado pela distribuição Multinomial.	31

Lista de tabelas

Tabela 1	– Distribuição para a variável resposta Y , e a natureza dos dados.	13
Tabela 2	– Distribuições de probabilidade da família exponencial e sua ligação canônica.	19
Tabela 3	– Respectivas médias e desvio padrão para a variável Tratamento (doses).	26
Tabela 4	– Respectivas médias e desvio padrão para a variável Avaliação.	27
Tabela 5	– Probabilidades estimadas para cada Nota (coluna) em cada tratamento (linha) segundo o modelo II.	28
Tabela 6	– Probabilidades estimadas para cada Nota (coluna) em cada Avaliação (linha) segundo o modelo II.	28
Tabela 7	– Estimativas e Erros-padrão (E.P.), Estatística de Wald e Razão de Probabilidades do modelo II ajustado.	29
Tabela 8	– Pseudo R^2	30

Sumário

1	INTRODUÇÃO	10
2	FUNDAMENTAÇÃO TEÓRICA	12
2.1	Modelos Lineares Generalizados	12
2.1.1	Distribuição Normal	14
2.1.2	Distribuição Poisson	14
2.1.3	Distribuição Binomial	14
2.1.4	Distribuição Gama	15
2.1.5	Distribuição Normal Inversa	16
2.1.6	Distribuição Multinomial	16
2.1.7	Distribuição Quasi-Poisson	17
2.1.8	Função Desvio	17
2.1.9	Estimação de β	18
2.1.10	Estimação de ϕ	19
2.1.11	Ligações Canônicas	19
2.1.12	Teste de Wald	19
2.1.13	Estatística de Pearson Generalizada	20
2.1.14	Método de Akaike	20
2.1.15	Técnicas de diagnóstico	21
2.1.16	Análise dos Resíduos	22
2.1.16.1	Resíduo de Pearson	23
2.1.16.2	Resíduo <i>Deviance</i>	23
3	MATERIAL E MÉTODOS	24
3.0.0.1	Modelo de Regressão Logística Multinomial	25
4	RESULTADOS E DISCUSSÃO	26
5	CONCLUSÃO	32
	REFERÊNCIAS	33
	Apêndices	35

1 Introdução

Dentre as oleaginosas cultivadas no Brasil o amendoim é um dos cultivares mais importantes para o agronegócio brasileiro. É uma planta originária da América do Sul, com provável centro de origem na região de Gran Chaco (Paraguai). A difusão do amendoim iniciou-se pelos indígenas para as diversas regiões da América Latina, América Central e México. No Brasil foi introduzido por volta do século XIX. A planta do amendoim é uma dicotiledônea da família Leguminosae, subfamília Papilionidae, gênero *Arachis*, que apresenta cerca de 80 espécies, amplamente distribuídas no bioma cerrado e em outros ambientes de vegetação aberta, tendo como limites de distribuição na Ilha de Marajó ao Norte, no Uruguai ao Sul, e por fim no Nordeste brasileiro a Leste e a Oeste, o sopé da Cordilheira dos Andes (GREGORY W.C., 1980).

Um dos principais fatores que afetam a sua produção são incidência de pragas, destaca-se o tripses do prateamento *Enneothrips flavens*, que é considerado a espécie de maior importância para a cultura (ROSSETO C.J., 1968);(BATISTA; GALLO; CARVALHO, 1973);(CASTRO, 1974); (RENSI AA. CALCAGNOLO, 1977). Os danos causados pelos tripses variam de acordo com a idade fisiológica da planta. As infestações do tripses do prateamento afetam todo desenvolvimento do amendoim com maiores danos durante o enchimento dos grãos. Esse é um inseto picador sugador que mede cerca de 2 mm de comprimento. As formas jovens apresentam coloração amarelada sem asas enquanto os adultos apresentam coloração escura e possuem asas franjadas. Esses insetos ficam abrigados em folíolos fechados ou semi-abertos onde perfuram o tecido para sugar a seiva que exsuda. São ovíparos e os ovos são colocados endofiticamente nos folíolos (NAKANO OCTAVIO SILVEIRA NETO, 1981);(GALLO D NAKANO, 2002). O ciclo evolutivo do tripses é de 13 dias, aproximadamente, passando pelas fases de ovo, ninfa, pré-pupa e pupa, desta surgindo o adulto. Os estágios ninfais e adulto ocorrem nas folhas e os demais, até a fase adulta, ocorrem no solo, geralmente nas fendas.

Ao contrário de outros insetos, tanto a pré-pupa como a pupa podem locomover-se quando molestadas (GALLO D NAKANO, 2002). Os adultos e as ninfas da maioria dos tripses apresentam uma tendência a alimentarem-se de forma agrupada causando danos mais pronunciados. De modo geral, as ninfas causam maiores danos pela alimentação do que os adultos, devido ao seu grande número, como por se alimentarem de forma mais agregada, uma vez que são menos ativas e restringem sua alimentação a áreas limitadas. Esses insetos raspam os folíolos. Com isso, aparecem manchas prateadas nas folhas, estrias e deformações. O ataque pode causar a redução de 10% a 70% na produção nas variedades de porte ereto, em geral, o ataque é maior em períodos quentes e secos.

O elemento potássio (*K*) está envolvido nos mecanismos de defesa das plantas a

pragas e doenças. As plantas bem nutridas em potássio apresentam redução na incidência, severidade e danos causados por insetos. Isso ocorre porque as altas concentrações de K nos tecidos favorecem a síntese e o acúmulo de compostos fenólicos, os quais atuam como inibidores de insetos e fungos (MASCARENHAS HAA. PATRÍCIO, 1995). O potássio possui papel fundamental no metabolismo de carboidratos e de produtos fosfatados presentes nas plantas, pois este se encontra ligado á resistência dos vegetais ás pragas e patógenos (CALCAGNOLO; LEITE; GALLO, 1974).

Os modelos lineares generalizados têm sido utilizados nas áreas de astronomia, agronomia, agricultura, saúde e pesca entre outras. Onde apresentam-se como ferramentas poderosas na análise de dados cujo interesse é o estudo da relação entre uma variável resposta, medida em escala contínua, discreta e qualitativa em função das variáveis preditoras, tanto de natureza quantitativa e/ou qualitativa. Além disso Cordeiro e Neto (2006) apresentaram um método iterativo para estimar os parâmetros, e ainda, introduziram o conceito de desvio, sendo muito importante na avaliação da qualidade do ajuste dos modelos, colaborando assim, com o desenvolvimento dos resíduos e técnicas de diagnósticos.

Assim, este trabalho teve por objetivo avaliar os efeitos da adubação fosfatada no controle do tripses do prateamento e seus reflexos na produtividade da cultura de amendoim por meio do uso de modelos lineares generalizados. 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 seleção e comparação a análise de desvio.

2 Fundamentação Teórica

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

2.1 Modelos Lineares Generalizados

Ao longo dos anos os modelos normais lineares foram utilizados para tentar descrever a maioria dos fenômenos aleatórios, ou seja, mesmo que tais fenômenos estudados não apresentem resposta para a pressuposição de normalidade, utiliza-se técnicas para a transformação de dados sugerida afim de encontrar a normalidade dos dados. Assim (NELDER; WEDDERBURN, 1972) propuseram os Modelos Lineares Generalizados que é uma reflexiva generalização da regressão de mínimos quadrados ordinários, que busca investigar o estudo da relação entre uma variável resposta, em função de diferentes variáveis preditoras (quantitativas e/ou qualitativas).

O uso dos modelos lineares generalizados foi introduzida por Nelder e Wedderburn (1972) como um quadro geral para o tratamento de uma série de modelos estatísticos para dados normais e não normais, tais como regressão linear múltipla, análise de variância (ANOVA), regressão logística, regressão de Poisson e modelos log-lineares. Wedderburn (1972) mostrou que uma série de técnicas estatísticas, usualmente estudadas separadamente, podia ser formulada, de uma maneira unificada como uma classe de modelos de regressão. A essa teoria unificadora de modelagem estatística, uma extensão dos modelos clássicos de regressão, deram o nome de modelos lineares generalizados. Além disso Nelder e Wedderburn (1972) também propuseram um método iterativo para estimar os parâmetros e introduziram o conceito de desvio, sendo muito importante na qualidade dos ajustes dos modelos, colaborando assim, com o desenvolvimento dos resíduos e técnicas de diagnósticos.

Quando se tem uma única variável aleatória Y , associada a um conjunto de variáveis explanatórias X_1, \dots, X_p , pode-se usar os modelos lineares generalizados. A família exponencial de distribuições garante aos parâmetros uma estatística suficiente. Sejam Y_1, \dots, Y_n variáveis aleatórias independentes, cada uma com função densidade (ou de probabilidade) na família exponencial da forma

$$f(y_i; \theta_i, \phi) = \exp[\phi\{y_i\theta_i - b(\theta_i)\} + c(y_i, \psi)], \quad i = 1, \dots, n \quad (2.1)$$

em que, $b(\cdot)$ e $c(\cdot)$ são funções conhecidas para cada observação, $b(\theta_i)$ é o parâmetro natural ou canônico e $\phi - 1 > 0$ é o parâmetro de dispersão (precisão).

Os modelos lineares generalizados envolvem os três componentes:

1. Componente aleatório

As variáveis respostas Y_1, Y_2, \dots, Y_n são independentes e seguem uma distribuição pertencente à família exponencial linear, com médias μ_i e parâmetro de escala constante ϕ e independentes das p variáveis explicativas.

2. Componente sistemático

O componente sistemático é representado por:

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

sendo que,

$$\eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$$

quando, η_i é o preditor linear, portanto, a parte sistemática que pode ser utilizada para realizar possíveis previsões para a equação da reta estimada.

3. Função de ligação

A função de ligação é uma função que liga o componente aleatório ao componente sistemático, ou seja, relaciona a média ao preditor linear.

Assim,

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

Na ordem serão apresentadas as funções θ , ϕ , $b(\theta)$ e $c(y; \phi)$, das principais distribuições de probabilidade.

Na Tabela 1, são apresentadas algumas distribuições e a natureza das suas variáveis (CORDEIRO, 2016).

Tabela 1 – Distribuição para a variável resposta Y , e a natureza dos dados.

Distribuição	Tipos de dados
Binomial	Proporção
Binomial Negativa	Contagem
Poisson	Contagem
Normal	Contínuos
Normal Inversa	Contínuos Assimétricos
Gama	Contínuos Assimétricos

A seguir serão apresentadas algumas demonstrações adaptadas das distribuições da família exponencial McCullagh e Nelder (1989).

2.1.1 Distribuição Normal

Seja uma variável aleatória Y com distribuição Normal com média μ e variância σ^2 , ou seja, $Y \sim N(\mu, \sigma^2)$, a função de densidade é dada por

$$\begin{aligned}
 f(y : \mu, \sigma^2) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(y - \mu)^2\right\}; y \in \mathbb{R} \\
 &= \exp\left\{\log \frac{1}{2\pi\sigma^2} - \frac{(y - \mu)^2}{2\sigma^2}\right\} \\
 &= \exp\left\{\frac{1}{\sigma^2} \left(\frac{-y^2 + 2y\mu - \mu^2}{2}\right) + \log \frac{1}{\sqrt{2\pi\sigma^2}}\right\} \\
 &= \exp\left\{\frac{1}{\sigma^2} \left(y\mu - \frac{\mu^2}{2}\right) - \frac{1}{2} \log 2\pi\sigma^2 - \frac{y^2}{2\sigma^2}\right\} \\
 &= \exp\left\{\frac{1}{\sigma^2} \left(y\mu - \frac{\mu^2}{2}\right) - \frac{1}{2} \left[\frac{y^2}{2} + \log 2\pi\sigma^2\right]\right\}
 \end{aligned}$$

em que, para $\theta = \mu$, $b(\theta) = \theta^2/2 = \mu^2/2$, $\phi = \sigma^2$, e $c(y, \psi) = \frac{1}{2}\left\{\frac{y^2}{\sigma^2} + \log 2\pi\sigma^2\right\}$.

2.1.2 Distribuição Poisson

Seja variável aleatória Y com distribuição Poisson. $Y \sim Poisson(\mu)$, sua função de probabilidade é dada por

$$\begin{aligned}
 P(Y = y) &= \frac{e^{-\mu} \mu^y}{y!}, y = 0, 1, 2, \dots \\
 &= \exp\{-\mu + y \log \mu - \log(y!)\} \\
 &= \exp\{y \log \mu - \mu - \log(y!)\}
 \end{aligned}$$

em que, $\mu > 0$. Fazendo $\theta = \log(\mu)$, $b(\theta) = e^\theta$, $\phi = 1$ e $c(y, \psi) = -\log(y!)$.

2.1.3 Distribuição Binomial

Seja Y a proporção de sucessos em n ensaios independentes de Bernoulli. Assumindo que $Y \sim Binomial(n, \pi)$, a função de probabilidades de Y é apresentada da seguinte forma

$$\begin{aligned}
P(Y = y) &= \binom{n}{y} \pi^y (1 - \pi)^{n-y}; \quad y = 0, 1, \dots, n. \\
&= \exp \left\{ y \log \pi + \log(1 - \pi)^{n-y} + \log \binom{n}{y} \right\} \\
&= \exp \left\{ y \log \pi - (n - y) \log(1 - \pi) + \log \binom{n}{y} \right\} \\
&= \exp \left\{ y [\log \pi - \log(1 - \pi)] + n \log(1 - \pi) + \log \binom{n}{y} \right\} \\
&= \exp \left\{ y \log \left(\frac{\pi}{1 - \pi} \right) + n \log(1 - \pi) + \log \binom{n}{y} \right\}
\end{aligned}$$

em que, $\phi = 1$, $\theta = \log \left(\frac{\pi}{1 - \pi} \right)$, $b(\theta) = n \log(1 - \pi)$ e $c(y, \psi) = \log \binom{n}{y}$.

2.1.4 Distribuição Gama

Seja Y uma variável aleatória com média μ e coeficiente de variação $\phi^{-1/2}$ ou seja, $Y \sim \text{Gama}(\mu, \phi)$. A função densidade de Y é apresentada da seguinte forma:

$$\begin{aligned}
f(y; \mu, \phi) &= \frac{1}{\Gamma(\phi)} \left(\frac{\phi y}{\mu} \right)^\phi \exp \left\{ -\frac{\phi y}{\mu} \right\} d(\log y) \\
&= \exp \left\{ \phi \left(-\frac{y}{\mu} - \log(\mu) \right) - \log \Gamma(\phi) + \phi \log(\phi y) - \log y \right\} \\
&= \exp \left\{ \log y^{\phi-1} + \log \phi^\phi - \log \mu - \frac{\phi y}{\mu} - \log \Gamma(\phi) \right\} \\
&= \exp \left\{ (\phi - 1) \log \mu + \phi \log \phi - \phi \log \mu - \frac{\phi y}{\mu} - \log \Gamma(\phi) \right\} \\
&= \exp \left\{ \phi \left[-\frac{y}{\mu} - \log \mu \right] + \phi \log \phi + \phi \log y - \log y - \log \Gamma(\phi) \right\} \\
&= \exp \left\{ \phi \left(\frac{-y}{\mu} - \log y \right) + \phi \log \phi y - \log y - \log \Gamma(\phi) \right\}
\end{aligned}$$

em que, $\theta = -1/\mu$, $\phi = \frac{1}{\phi}2$, $b(\theta) = -\log(-\theta)$ e $c(y, \psi) = \phi \log \phi y - \log y - \log \Gamma(\phi)$.

2.1.5 Distribuição Normal Inversa

Seja Y uma variável aleatória com distribuição normal inversa de média μ e parâmetro de precisão ϕ , denotada por $Y \sim NI(\mu, \phi)$ e cuja função densidade é dada por

$$\begin{aligned} f(y; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2y^3}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2\mu^2y}\right\} \\ &= \exp\left\{\frac{-y^2 + 2y\mu - \mu^2}{2\sigma^2\mu^2y} + \log(2\pi\sigma^2y^3) - \frac{1}{2}\right\} \\ &= \exp\left\{\frac{1}{\sigma^2}\left[-y\frac{1}{2\mu^2} + \frac{1}{\mu}\right] - \frac{1}{2}\log(2\pi\sigma^2y^3) - \frac{1}{2\sigma^2y}\right\} \\ &= \exp\left\{\frac{\left(\frac{-y}{2\mu^2} + \frac{1}{\mu}\right)}{\sigma^2} - \frac{1}{2}\left[\log(2\pi\sigma^2y^3) - \frac{1}{\sigma^2y}\right]\right\} \end{aligned}$$

em que $\theta = -\frac{1}{2\mu^2}$, $\phi = \sigma^2$, $b(\theta) = \frac{1}{\mu}$ e $c(y, \psi) = -\frac{1}{2}\left[\log(2\pi\sigma^2y^3) - \frac{1}{\sigma^2y}\right]$.

2.1.6 Distribuição Multinomial

Seja Y um vetor aleatório $Y = (\mathbf{y}_1, \dots, \mathbf{y}_k)$ que segue uma distribuição multinomial com parâmetros n e π onde $\pi = (\pi_1, \dots, \pi_k)$. A função de probabilidade de Y é apresentada da seguinte forma,

$$\begin{aligned} P(Y = y) &= \frac{M!}{y_1! \times \dots \times y_k!} \times \pi_1^{y_1} \times \pi_2^{y_2} \times \dots \times \pi_k^{y_k} \\ &= \frac{M!}{y_1! \times \dots \times y_k!} \exp\left\{\sum_{k=1}^K y_k \log \pi_k\right\}. \end{aligned}$$

Sendo que $\sum_{k=1}^K x_k = M$, para conseguir uma representação mínima para o multinomial, usamos os primeiros componentes $k-1$ de π :

$$\begin{aligned} P(Y = y) &= \exp\left\{\sum_{k=1}^{K-1} \log(\pi_k)y_k + \log\left(1 - \sum_{k=1}^{k-1} \pi_k\right)\left(n - \sum_{k=1}^{k-1} \pi_k\right)\right\} \frac{M!}{y_1!y_2! \dots y_k!} \\ &= \exp\left\{\sum_{k=1}^{k-1} y_k \log\left(\frac{\pi_k}{1 - \sum_{k=1}^{k-1} \pi_k}\right) + n \log\left(1 - \sum_{k=1}^{k-1} \pi_k\right)\right\} \frac{M!}{y_1!y_2! \dots y_k!} \end{aligned}$$

Para o qual usamos o fato de que $\pi_k = 1 - \sum_{k=1}^{K-1} \pi_j$. A partir desta representação, obtemos:

$$\eta_k = \log\left(\frac{\pi_k}{1 - \sum_{k=1}^{K-1} \pi_j}\right) = \log\left(\frac{\pi_k}{\pi_j}\right),$$

em que, $\theta = \frac{\pi_k}{1 - \sum_{k=1}^{k-1} \pi_j}$, $\phi = 1$, $b(\theta) = \log(1 - \sum_{k=1}^{k-1} \pi_j)$, $c(y, \psi) = \log\left(\sum_{k=1}^{k-1} e^{\eta_k}\right)$.

2.1.7 Distribuição Quasi-Poisson

Para modelo Quasi-Poisson, do qual, pressupondo que $Y \sim P(\lambda)$ cuja a função de probabilidade é dada por

$$\Pr(Y = y) = \frac{e^{-\lambda} - \lambda^y}{y!}, \quad y = 0, 1, 2, \dots$$

Pode-se mostrar que quando $\lambda \rightarrow \infty$

$$(y - \lambda)/\sqrt{\lambda} \xrightarrow{d} N(0, 1).$$

Para λ grande temos que Y segue aproximadamente uma distribuição normal de média λ e desvio padrão $\sqrt{\lambda}$. Ainda assim, se queremos aplicar um modelo normal linear para explicar λ . Segundo (MACIEL, 2016) teremos o inconveniente do desvio padrão depender da média, o que inviabiliza o uso de um modelo normal linear homocedástico. Uma maneira de contornarmos esse problema é através da aplicação de uma transformação na resposta Y de maneira a alcançarmos a normalidade e a constância de variância, mesmo que aproximadamente. Assim, se Y é Poisson, segue que quando $\lambda \rightarrow \infty$ tem-se

$$\{\sqrt{Y} - E(\sqrt{Y})\} \xrightarrow{d} N(0, 1/4).$$

Então, quando λ é grande, a variável aleatória $2\{\sqrt{Y} - E(\sqrt{Y})\}$ segue aproximadamente uma distribuição $N(0, 1)$. Desta forma, se temos uma amostra aleatória Y_1, \dots, Y_n visto que $Y_i \sim P(\lambda_i)$ e queremos explicar λ_i através de variáveis explicativas, podemos propor para λ_i grande, o modelo normal linear abaixo

$$\sqrt{Y_i} = X_i^T \beta + \varepsilon_i,$$

em que, $\varepsilon_i \sim N(0, \sigma^2), i = 1, \dots, n$.

2.1.8 Função Desvio

Um dos objetivos da função desvio é analisar a adequação do modelo como um todo e a realização de uma investigação minuciosa quanto às discrepâncias locais que, no caso de serem significativas, podem levar a escolha de um outro modelo, diferente do inicialmente proposto.

Suponhamos que o logaritmo da função de verossimilhança seja definido por,

$$L(\hat{\mu}; \mathbf{y}) = \sum_{i=1}^n L(\hat{\mu}; \mathbf{y}_i),$$

em que $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = x_i^T \beta$. $L(\hat{\mu}; \mathbf{y})$ é o máximo da log-verossimilhança para o modelo em estudo com p parâmetros, $L(\mathbf{y}; \mathbf{y})$ para o modelo saturado com n parâmetros.

A qualidade do ajuste de um MLG é avaliada através da função desvio, escrita da seguinte forma McCullagh e Nelder (1989):

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \phi D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\{L(\mathbf{y}; \mathbf{y}) - L(\hat{\boldsymbol{\mu}}; \mathbf{y})\}, \quad (2.2)$$

onde

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n [y_i(\tilde{\theta}_i - \hat{\theta}_i) + (b(\tilde{\theta}_i) - b(\hat{\theta}_i))]. \quad (2.3)$$

em que $\tilde{\theta}_i = \theta_i(\hat{\mu}_i)$ e $\hat{\theta}_i = \theta_i(\tilde{\mu}_i)$, respectivamente, as estimativas de máxima verossimilhança de θ_i para os modelos com p parâmetros ($p < n$) e saturado ($p = n$). $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ tem, assintoticamente, distribuição χ^2 com $n - p$ graus de liberdade, sendo n o número de observações e p , o número de parâmetros do modelo.

2.1.9 Estimação de $\boldsymbol{\beta}$

A obtenção da estimativa de máxima verossimilhança de $\boldsymbol{\beta}$ através do processo iterativo de Newton-Raphson é determinada pela expansão da função escore $\mathbf{U}(\boldsymbol{\beta})$ em torno de um valor inicial $\boldsymbol{\beta}^{(0)}$ (NELDER; WEDDERBURN, 1972), em que

$$\mathbf{U}(\boldsymbol{\beta}) \cong \mathbf{U}(\boldsymbol{\beta}^{(0)}) + \mathbf{U}'(\boldsymbol{\beta}^{(0)})(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)})$$

Onde $\mathbf{U}'(\boldsymbol{\beta})$ será a derivada primeira de $\mathbf{U}_{\boldsymbol{\beta}}$ em relação a $\boldsymbol{\beta}$. Dessa forma, seguindo o método acima, chega-se ao processo iterativo que é dado por

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + [(-\mathbf{U}'_{\boldsymbol{\beta}})^{-1}]^{(m)} \mathbf{U}_{\boldsymbol{\beta}}^{(m)}, \text{ com } m = 0, 1, \dots, \quad (2.4)$$

A aplicação do método de escore de Fisher substituindo a matriz $-\mathbf{U}_{\boldsymbol{\beta}}$ pelo correspondente valor esperado pode ser mais conveniente quando não se sabe se a matriz $-\mathbf{U}(\boldsymbol{\beta})$ é positiva definida (BARROS, 2011). Isso resulta no seguinte processo iterativo:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + [\mathbf{K}_{\boldsymbol{\beta}}^{-1}]^{(m)} \mathbf{U}_{\boldsymbol{\beta}}^{(m)}, \text{ com } m = 0, 1, \dots,$$

Este processo pode ser reescrito como um processo iterativo de mínimos quadrados ponderados

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)} \quad (2.5)$$

com $\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{-\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{y} - \boldsymbol{\mu})$. A quantidade \mathbf{z} desempenha o papel de uma variável dependente modificada, enquanto \mathbf{W} é uma matriz de pesos que muda a cada passo do processo iterativo.

2.1.10 Estimação de ϕ

Segundo Cordeiro e Neto (2006), igualando a função escore $U_{(\phi)}$ a zero tem-se

$$\sum_{i=1}^n c'(y_i, \hat{\phi}) = \frac{1}{2} D(\mathbf{y}; \hat{\boldsymbol{\mu}}) - \sum_{i=1}^n \{y_i \tilde{\theta}_i - b(\tilde{\theta}_i)\}$$

no qual, $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ mostra o desvio do modelo de interesse. De acordo com o modelo que estiver sendo trabalhado é necessário trabalhar com a expressão acima para chegar a um estimador para o parâmetro ϕ .

2.1.11 Ligações Canônicas

Uma particularidade relevante ocorre quando o parâmetro canônico θ coincide com o preditor linear, obtêm-se uma ligação canônica, isto é, quando

$$\theta_i = \eta_i$$

Neste caso,

$$L(\beta) = g^{-1}(\eta_i) = g^{-1}(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}) \quad (2.6)$$

Um dos benefícios de usarmos ligações canônicas é que as mesmas garantem a concavidade de $L(\beta)$ e conseqüentemente muitos resultados assintóticos são obtidos mais facilmente (CORDEIRO; NETO, 2006). As ligações canônicas para os modelos mais usuais estão apresentados na Tabela 2.

Tabela 2 – Distribuições de probabilidade da família exponencial e sua ligação canônica.

Distribuição	η
Normal	μ
Poisson	$\log \mu$
Binomial	$\log \frac{\mu}{(1-\mu)}$
Gama	$\frac{1}{\mu}$
Normal Inversa	$\frac{1}{\mu^2}$

2.1.12 Teste de Wald

Suponhamos que pretende-se testar a hipótese nula

$$H_0 : \beta_j = 0, j = 0, \dots, p,$$

que indica que o coeficiente independente $\beta_0 = 0$ é irrelevante para o modelo ($j = 0$) ou que a variável explicativa X_j não deve constar do modelo de regressão ($j \neq 0$). A

estatística de Wald e expressa da seguinte forma

$$W = [\hat{\beta} - \beta^0]^T \hat{Var}^{-1}(\hat{\beta})[\hat{\beta} - \beta^0],$$

em que $\hat{Var}(\hat{\beta})$ denota a matriz de variância-covariância assintótica de $\hat{\beta}$ estimada em $\hat{\beta}$. Usando os resultados conhecidos de álgebra linear, mostramos que a variância assintótica de $\hat{\beta}$ é dada por

$$Var(\hat{\beta}) = \phi^{-1} [X_1^T W^{1/2} M_2 W^{1/2} X_1]^{-1},$$

em que X_1 sai da partição $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, sendo portanto $n \times q$, \mathbf{X}_2 é $n \times (p - q)$, $\mathbf{M}_2 = \mathbf{I}_n - \mathbf{H}_2$ e $\mathbf{H}_2 = \mathbf{W}^{1/2} \mathbf{X}_2 (\mathbf{X}_2^T \mathbf{W} \mathbf{X}_2)^{-1} \mathbf{X}_2^T \mathbf{W}^{1/2}$ e a matriz de projeção ortogonal de vetores do R^n no subespaço gerado pelas colunas da matriz $\mathbf{W}^{1/2} \mathbf{X}_2$. Em particular, no caso normal linear, temos as simplificações $\mathbf{H}_2 = \mathbf{X}_2 (\mathbf{X}_2^T \mathbf{X}_2)^{-1} \mathbf{X}_2^T$ e $Var(\hat{\beta}_1) = \sigma^2 [X_1^T (\mathbf{I}_n - \mathbf{H}_2) X_1]^{-1}$.

Em particular, para o caso de $p = 1$, o teste de Wald baseado na distribuição normal assintótica de β é uma generalização da estatística t de Student (WALD, 1943).

$$W = \frac{(\hat{\beta} - \beta^0)^2}{\hat{Var}(\hat{\beta})}.$$

Um problema com a estatística de Wald, especialmente quando $\boldsymbol{\eta}(\boldsymbol{\beta})$ é não linear em $\boldsymbol{\beta}$, é a dependência de W com a parametrização utilizada. Isto é, duas formas diferentes e equivalentes para $\boldsymbol{\eta}(\boldsymbol{\beta})$, podem levar a diferentes valores de W .

2.1.13 Estatística de Pearson Generalizada

Uma importante medida de discrepância do modelo ajustado em relação aos dados é a estatística de Pearson generalizada definida por

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

em que, $V(\hat{\mu}_i)$ é a função de variância estimada para a distribuição proposta para os dados.

Tanto a *função desvio* quanto a *estatística de Pearson generalizada* têm, para o modelo normal linear, distribuição χ^2 exata. Resultados assintóticos são possíveis para outras distribuições. Segundo Cordeiro e Neto (2006) a vantagem da função desvio é que ela é aditiva e acrescentando-se variáveis explicativas ao modelo, o desvio deve decrescer, diferentemente de χ^2 : Contudo, χ^2 é algumas vezes preferível pois tem uma interpretação simples.

2.1.14 Método de Akaike

Este método realiza um processo de minimização que não envolve testes estatísticos. A ideia básica é selecionarmos um modelo que seja parcimonioso, ou em outras palavras,

que esteja bem ajustado e tenha um número reduzido de parâmetros. Segundo (AKAIKE, 1974) o logaritmo da função de verossimilhança cresce com o aumento do número de parâmetros do modelo, uma proposta seria encontrarmos o modelo com menor valor para a função

$$AIC = -L(\hat{\beta}) + p,$$

em que p denota o número de parâmetros.

No caso do modelo normal linear é possível mostrar que AIC fica expresso, quando σ^2 é desconhecido, na forma

$$AIC = n \log \{D(\mathbf{y}; \hat{\boldsymbol{\mu}})/n\} + 2n,$$

em que, $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2$.

2.1.15 Técnicas de diagnóstico

Essas técnicas são de grande importância para a validação de um modelo pois se as suas pressuposições forem violadas, os resultados não serão confiáveis, assim a previsão do modelo estará comprometida (MYERS RAYMOND H. MONTGOMERY, 2012). Existem algumas formas de avaliar a qualidade do ajuste dos modelos aos dados, levando a selecionar através de métodos qual o melhor modelo que representa a relação das variáveis explicativas com a variável resposta, assim tais relações são:

- i) Técnicas de diagnóstico: detectar pontos de alavanca, avaliar os resíduos, investigar pontos de influência e outras técnicas gráficas;
- ii) Métodos de seleção de modelos: método normal linear, Critério de Akaike(AIC), entre outros.

Segundo Cordeiro e Neto (2006), existem três tipos de problemas de ajuste nos modelos, o primeiro é a presença de pontos mal ajustados, no caso pontos aberrantes; o segundo problema é a violação dos pressupostos para os erros e ou para as estruturas das médias; e por último, é a presença de observações influentes.

Um procedimento informal para verificar a função de ligação consiste na construção de um gráfico entre a variável dependente ajustada e o preditor linear. Então se os dados plotados no gráfico forem aproximadamente lineares, a função de ligação estará correta. Para dados binários este gráfico é não informativo, sendo necessário o uso de métodos formais.

As técnicas gráficas mais recomendadas para os MLGs são:

- i) Gráficos dos resíduos contra os valores ajustados: pode mostrar a existência de valores discrepantes como também heterogeneidade de variância. O padrão do gráfico deve ser uma distribuição dos resíduos em torno de zero com amplitude constante;
- ii) Gráfico dos resíduos contra as variáveis explicativas: pode mostrar se existe uma relação sistemática (relação não linear, não normalidade, heterocedasticidade, etc) entre os resíduos e uma variável explicativa. O padrão do gráfico deve ser uma distribuição dos resíduos em torno de zero com amplitude constante;
- iii) Gráfico dos resíduos contra a ordem das observações: pode auxiliar na detecção de alguma variável altamente correlacionada com a sequência do tempo (ou espaço) que as observações foram coletadas;
- iv) Gráficos da variável dependente ajustada contra o preditor linear: ajuda a verificar a adequação da função e o padrão esperado e uma tendência linear, indicando a adequação da função de ligação;
- v) Gráficos dos resíduos absolutos contra valores ajustados contra valores ajustados: serve para identificar se a função de variância adotada é adequada, o comportamento esperado dos pontos é uma distribuição aleatória em torno de zero e amplitude constante.
- vi) Gráfico normal e meio normal de probabilidades (*normal plot e half-normal plot*): são úteis para verificar se a função de variância foi corretamente especificada, e detectar a presença *outlier*. o comportamento esperado dos resíduos para um adequado é aproximadamente uma reta.

2.1.16 Análise dos Resíduos

A análise dos resíduos para os Modelos Lineares Generalizados é semelhante as usa das em modelos clássicos, porém com algumas adaptações. Para verificar a pressuposição de linearidade para o modelo clássico utiliza-se vetores \mathbf{Y} e $\hat{\mu}$, enquanto os MLGs utiliza-se a variável dependente ajustada \mathbf{z} e o preditor linear $\hat{\eta}$. Segundo McCullagh e Nelder (1989) a variância residual é substituída por uma estimativa de ϕ e a matriz de projeção \mathbf{H} ou matriz "hat", é definida por: $\mathbf{W}^{\frac{1}{2}}\mathbf{X}\mathbf{I}^{-1}(\beta)\mathbf{X}^T\mathbf{W}^{\frac{1}{2}}$

Os resíduos da modelagem estatística têm um papel muito importante que está relacionada com a qualidade do ajuste, constituindo uma das etapas mais importantes no processo de escolha do modelo adequado. Nos MLGs, segundo Cordeiro e Neto (2006), os resíduos são usados para explorar a adequação do modelo ajustado com respeito à escolha da função de variância, da função de ligação e de termos no preditor linear. Além disso, eles também são úteis na identificação de pontos aberrantes, que poderão ser influentes

ou não. Os resíduos medem discrepâncias entre os valores observados y_i e seus valores ajustados $\hat{\mu}_i$.

2.1.16.1 Resíduo de Pearson

O resíduo de Pearson tem a seguinte expressão:

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}} \quad (2.7)$$

Para os modelos log-lineares tem-se que $r_i^P = (y_i - \hat{\mu}_i)\hat{\mu}_i^{-\frac{1}{2}}$ e (HABERMAN, 1974) sugere a correção r_i^{P*} , em que $\mathbf{Z} = \{z_{ij}\} = \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T$ com $\mathbf{W} = \text{diag}\{\mu_i\}$, para tornar a distribuição do resíduo r_i^P , aproximadamente, normal $N(0, 1)$. A variância média dos r_i^{P*} é $1 - (1+p)/n$. Podem-se incorporar pesos a priori na fórmula (2.6). A desvantagem deste resíduo é que sua distribuição é, geralmente, bastante assimétrica para modelos não-normais (CORDEIRO; NETO, 2006).

2.1.16.2 Resíduo *Deviance*

Como nos MLG a *deviance* é usada como medida de discrepância, então cada unidade i contribui com uma quantidade d_i (MCCULLAGH; NELDER, 1989), de tal modo que

$$\sum_{i=1}^n d_i = \frac{2w_i}{\phi} [y_i(\tilde{\theta}_i - \hat{\theta}_i) + (b(\tilde{\theta}_i) - b(\hat{\theta}_i))].$$

Sendo $\tilde{\theta}_i$ e $\hat{\theta}_i$ as estimativas do parâmetro canônico sob os modelos saturados. Define-se então o resíduo *deviance* correspondente a cada resposta:

$$r_{Di} = \pm (y_i, \hat{\mu}_i) \sqrt{d_i} \quad (2.8)$$

Conforme pode-se verificar a diferença do valor real e o valor estimado, e d_i a respectiva função de desvio para o modelo. Neste tipo de resíduos é também possível obter a função estandardizada através da multiplicação de r_{Di} por $1/\sqrt{\hat{\phi}(1-h_i)}$.

$$r_{Di}^* = \frac{r_{Di}}{\sqrt{\hat{\phi}(1-h_i)}} \quad (2.9)$$

3 Material e Métodos

Os dados utilizado neste estudo são provenientes de um experimento cujo delineamento o foi de blocos casualizados no esquema de parcelas subdivididas no tempo, para avaliar como tratamentos as doses de 0, 50, 100, 150 e 200 *kg/ha* de potássio, aplicado em cobertura, 8 dias após emergência das plantas, com 5 repetições. Cada unidade experimental foi constituída de 5 linhas de 4 m de comprimento espaçadas de 0,90 m entre linhas. Na semeadura, realizada em 14 de dezembro de 2012, utilizou-se o cultivar de amendoim IAC 886, na densidade de 15 sementes m^2 e densidade final de 12 plantas m^2 . Nas avaliações, realizadas a partir dos 10 dias após emergência das plantas e as subsequentes a cada 10 dias, foram coletados, ao acaso, 10 folíolos fechados ou semiabertos por parcela, do terço superior das plantas, nas duas linhas centrais, para contagem de adultos e ninfas do tripses.

Uma escala de notas variando de 1 a 5 foi utilizada para quantificar os sintomas de injúrias causadas pelos tripses nas plantas de amendoim. A escala de notas proposta por (MORAES, 2005) que recomendam atribuir visualmente as plantas notas que variam de 1 a 5, caracterizadas de acordo com os sintomas:

- a) Nota 1 - folíolos com ausência de sintomas;
- b) Nota 2 - folíolos com poucas pontuações prateadas, sem deformações;
- c) Nota 3 - folíolos com poucas pontuações prateadas, com início de enrolamento das bordas dos folíolos;
- d) Nota 4 - folíolos com pontuações prateadas generalizadas, com enrolamento das bordas;
- e) Nota 5 - folíolos com pontuações prateadas generalizadas, com encarquilhamento total desses folíolos.

O modelo utilizado para avaliar o comportamento dos dados segue uma Distribuição Multinomial. Considera-se que Y_i uma variável resposta que pode assumir valores discretos, nomeadamente, $1, 2, \dots, j$.

$$p_{ij} = P(Y_i = j)$$

A distribuição de probabilidade para a variável Y_{ij} dado o total de casos n_i é dado por:

$$P(Y_{i1} = y_{i1}, \dots, Y_{ij} = y_{ij}) = \binom{n_i}{y_{i1}, \dots, y_{ij}} p_{1i}^{y_{i1}} \times \dots \times p_{ij}^{y_{ij}}$$

3.0.0.1 Modelo de Regressão Logística Multinomial

Para a variável resposta multinomial temos $j-1$ regressões logísticas. Genericamente, podemos definir o modelo logístico multinomial da seguinte forma:

$$\log\left(\frac{p_{ij}}{p_{iJ}}\right) = \alpha_j + x_i'\beta_j \quad j = 1, \dots, J-1, \quad i = 1, 2, \dots, n \quad (3.1)$$

em que x_i é o vetor de covariáveis, α_j são os interceptos e β_j é o vetor de coeficientes de regressão. Os modelos de regressão logística multinomial também são conhecidos como modelos logit quando os preditores são categóricos.

Para avaliar a qualidade do ajuste do modelo de regressão logística multinomial, será utilizado neste estudo o Pseudo R^2 , que explica o poder total das variáveis independentes do modelo.

O ajuste dos modelos foi feito por meio do *software* gratuito R (R Core Team, 2017). Realizou-se a visualização gráfica, com o intuito de identificar a qualidade do ajuste do modelo, para tanto, foram construídos gráficos semi-normais de probabilidade com envelope simulado para os componentes do desvio, usando-se o pacote *hnp* (MORAL, 2013), com nível de confiança aproximado de 0,95.

Essa técnica foi proposta por Atkinson (1985) para verificarmos a adequação do ajuste de um determinado modelo a um conjunto de observações e utilizaremos o gráfico de probabilidade meio-normal com envelope simulado. Que consiste em se plotar os valores absolutos ordenados de uma determinada medida de diagnóstico (diferentes tipos de resíduos, distância de Cook, entre outros.) *versus* as estatísticas de ordem esperadas da distribuição meio-normal, que calculamos por

$$\Phi^{-1}\left(\frac{i + n - \frac{1}{8}}{2n + \frac{1}{2}}\right)$$

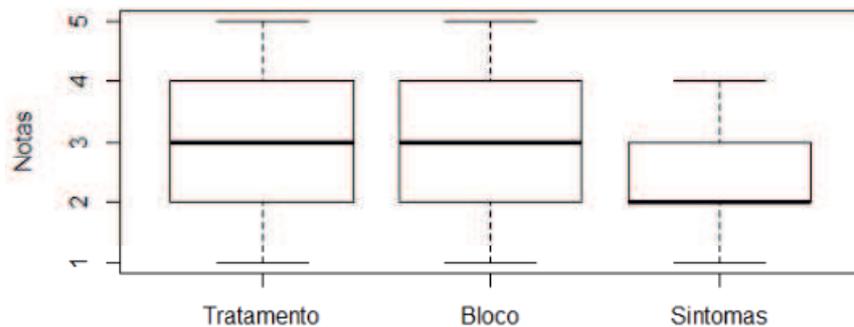
Sendo n o tamanho da amostra e $i = 1, \dots, n$, adicionando-se a esse gráfico um envelope simulado. Esse gráfico pode ser utilizado mesmo que os resíduos não tenham distribuição de probabilidade normal (NETER J KUTNER, 1996).

4 Resultados e Discussão

Nesta seção são apresentados e discutidos os resultados obtidos a partir da análise do conjunto de dados. Todos foram obtidos a partir do *software* estatístico R (R Core Team, 2017) e os comandos encontra-se no apêndices.

Inicialmente é apresentada na Figura 1 os fatores tratamento, bloco e sintomas utilizados no fenômeno estudado a partir do gráfico boxplot, porém, só vamos explorar a variável resposta sintomas. Nela observa-se que sua amplitude vai de 1 à 4, com uma assimetria à direita, ou seja, maior concentração nas primeiras Notas.

Figura 1 – Boxplot dos dados da cultura do Amendoim.



Obtivemos algumas estatísticas descritivas para as notas em função da variável tratamento na Tabela 3 e suas respectivas médias das Notas e desvio-padrão. De forma geral a média das Notas dos sintomas para cada tratamento foram baixas, em torno de 2, com apenas os tratamentos 3 e 4, obtendo médias abaixo de 2.

Tabela 3 – Respectivas médias e desvio padrão para a variável Tratamento (doses).

Tratamentos	Média	Desvio-padrão
0	2,20	0,8081220
50	2,04	0,8320126
100	1,94	0,5858885
150	1,74	0,5646021
200	2,22	0,7637181

Na Tabela 4 apresentam-se as médias das Notas para variável avaliação (AVAL), ou seja, os dias de avaliação, percebe-se que até o quinto dia as Notas de avaliação vão

aumentando, no entanto a partir do quinto dia houve um decréscimo das Notas médias dos sintomas.

Tabela 4 – Respectivas médias e desvio padrão para a variável Avaliação.

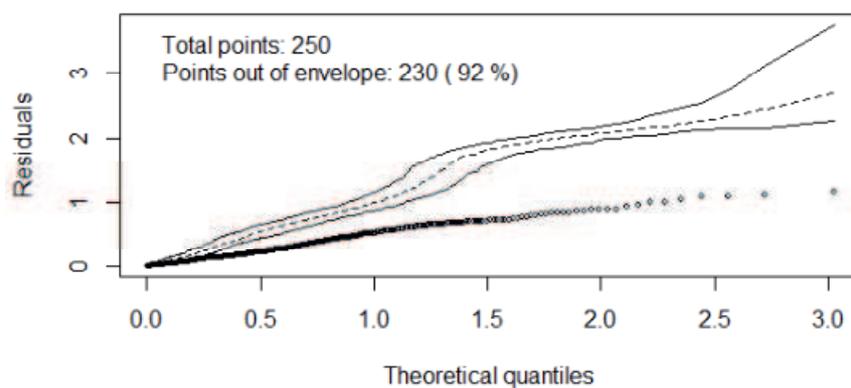
Avaliações	Média	Desvio-Padrão
1	1,08	0,2768875
2	1,44	0,5066228
3	1,80	0,4082483
4	2,32	0,6904105
5	2,56	0,5066228
6	2,04	0,4546061
7	2,32	0,7483315
8	2,44	0,8698659
9	2,28	0,6137318
10	2,00	0,6454972

Tanto na Tabela 3 como na Tabela 4 os desvios-padrão são pequenos, menores que uma unidade, ou seja, há indícios que os valores dos estão próximos da média.

Foram ajustados dois modelos, modelo I considerado completo pois tinha todos os efeitos principais (Avaliação e Tratamento) e o efeito de interação (Avaliação \times Tratamento), porém este foi ajustado sobre a suposição que os dados seguissem uma Distribuição de Poisson e o modelo II com os mesmos efeitos mas sob distribuição Multinomial.

O modelo I foi descartado pois o gráfico normal de probabilidades com envelope simulado `hnp`, obteve mais de 92% dos pontos fora do envelope, evidenciando um ajuste ruim.

Figura 2 – Modelo ajustado pela distribuição Poisson.



Além do fato que, por se tratar de notas a característica da variável não se enquadrava propriamente dentro de uma Distribuição Poisson, que é caracterizada por

dados de contagem. Para confirmar tais fatos, foi utilizado o pacote `vcd`, em que, procedeu-se com o teste de adequabilidade do modelo (bondade de ajuste), sob hipótese que a distribuição ajustada é adequada. O resultado encontrado foi de $3,184358 \times 10^{-25}$ portanto, valor $P < 0,05$, indica que a Distribuição Poisson não se adequa aos dados analisados. Foi utilizado ainda a desvio residual do modelo em conjunto com os graus de liberdade, para verificar a adequação do modelo linear generalizado sob ajuste da Distribuição Poisson. O resultado foi um valor $P < 0,05$, indicando que o modelo não se ajusta bem aos dados.

O modelo II foi ajustado e validado pela distribuição Multinomial e foram estimadas as probabilidades marginais de cada Nota por tratamento apresentadas na Tabela 5.

Tabela 5 – Probabilidades estimadas para cada Nota (coluna) em cada tratamento (linha) segundo o modelo II.

Tratamentos/Nota	1	2	3	4
0	0,19	0,52	0,26	0,03
50	0,19	0,54	0,26	0,01
100	0,20	0,55	0,25	0,00
150	0,20	0,56	0,24	0,00
200	0,20	0,56	0,24	0,00

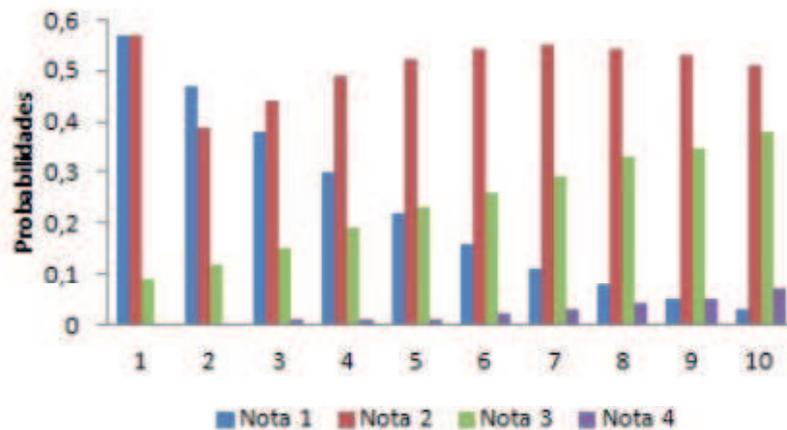
De acordo com a Tabela 5, nota-se que a maior probabilidade está em torno da nota 2, para qualquer tratamento avaliado, com divergências além do terceiro dígito significativo nas probabilidades. Deste modo, todos os tratamentos apresentam probabilidades parecidas de produzirem as mesmas notas.

Na Tabela 6 e Figura 3, tem-se as probabilidades estimadas para os dias de avaliação, percebe-se que os primeiros 5 dias tem probabilidade maior de computar Nota 1 para os sintomas de prateamento na planta, sendo maiores estas probabilidades para os dois primeiros dias de avaliação, as Notas 2 aparece com maior probabilidade entre os dias de avaliação 3 à 10, respectivamente. Este fato é comprovado via as estatísticas e significâncias encontradas na Tabela 7.

Tabela 6 – Probabilidades estimadas para cada Nota (coluna) em cada Avaliação (linha) segundo o modelo II.

Dias de Avaliação/Nota	1	2	3	4
1	0,57	0,33	0,09	0,00
2	0,47	0,39	0,12	0,00
3	0,38	0,44	0,15	0,01
4	0,30	0,49	0,19	0,01
5	0,22	0,52	0,23	0,01
6	0,16	0,54	0,26	0,02
7	0,11	0,55	0,29	0,03
8	0,08	0,54	0,33	0,04
9	0,05	0,53	0,35	0,05
10	0,03	0,51	0,38	0,07

Figura 3 – Gráfico de probabilidade ajustado para os dias de Avaliação.



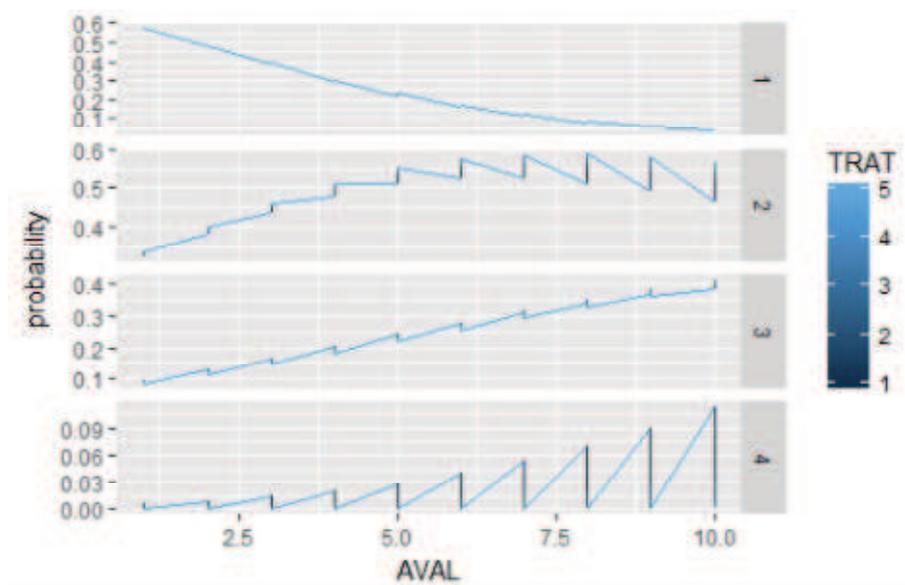
O modelo II (Tabela 7), tal como o modelo I, foi ajustado com os efeitos principais e de interação. As avaliações surtiram efeito significativo (valor $P < 0,05$) no modelo ajustado com a distribuição multinomial, e os tratamentos apresentaram valor $P > 0,05$. O efeito da interação via estatística de Wald foi significativa conforme comprovado com o cálculo dos valores P para cada Nota.

Tabela 7 – Estimativas e Erros-padrão (E.P.), Estatística de Wald e Razão de Probabilidades do modelo II ajustado.

Notas	Intercepto (E.P.)	Bloco (E.P.)	Tratamento (E.P.)	Avaliação (E.P.)	Trat×Aval (E.P.)
2	-1,29 (0,87)	0,04 (0,11)	0,06 (0,23)	0,45 (0,17)	-0,03(0,04)
3	-4,54 (1,28)	-0,11 (0,14)	0,78 (0,33)	0,94 (0,21)	-0,14 (0,05)
4	-3,94 (4,59)	-0,95 (0,64)	-0,31 (2,50)	1,07 (0,61)	-0,19 (0,34)
Wald (Valor P)	Intercepto (Pr.)	Bloco (Pr.)	Tratamento (Pr.)	Avaliação (Pr.)	Trat×Aval (Pr.)
2	-1,48 (0,13)	0,39 (0,69)	0,28 (0,77)	2,57 (0,01)	-0,61 (0,54)
3	-3,54 (0,00)	-0,74 (0,46)	2,36 (0,01)	4,33 (0,00)	-2,50 (0,01)
4	-0,86 (0,39)	-1,48 (0,13)	-0,13 (0,89)	1,75 (0,08)	-0,54 (0,58)
Razão (Pr.)	Intercepto ($e(\beta_j)$)	Bloco ($e(\beta_j)$)	Tratamento ($e(\beta_j)$)	Avaliação ($e(\beta_j)$)	Trat×Aval ($e(\beta_j)$)
2	0,27	1,05	1,07	1,58*	0,97
3	0,01	0,90	2,18	2,56	0,86
4	0,02	0,39	0,73	2,91	0,83

De acordo com a tabela 7, há 14% menos chance de se aumentar da Nota 1 para à Nota 3, a medida que se variam os efeitos de Tratamento e Avaliação. Há 58% mais chance de se mudar a Nota de 1 para 2 e quase 3* vezes mais chance de se mudar a Nota de 1 para 4, quando se muda os dias de Avaliação.

Figura 4 – Gráfico de risco ajustado pela distribuição Multinomial.



Conforme Figura 4 procurou-se entender o modelo usando probabilidades previstas, tais probabilidades estão associadas aos sintomas na planta e dias de avaliações. O risco relativo para um aumento de uma unidade na variável avaliação é de 1,42, isto indica que a medida que se aumenta os dias de avaliação aumenta-se a probabilidade de se obter as notas 2, 3 e 4.

Para modelos estatísticos baseados em estimativas da máxima verossimilhança, como um MLG, é mais conveniente construir uma estimativa do R^2 , chamada de *pseudo* – R^2 . Seu valor está no intervalo de 0 a 1: Quanto maior, mais explicativo é o modelo. Segundo Brito e Neto (2008), a avaliação do nível de um modelo logístico pode ser realizada por meio do *Likelihood ou probabilidade* dos *pseudo* R^2 , é uma medida da qualidade geral do modelo e quanto maior seu valor, melhor é o ajuste do modelo como um todo.

Tabela 8 – Pseudo R^2 .

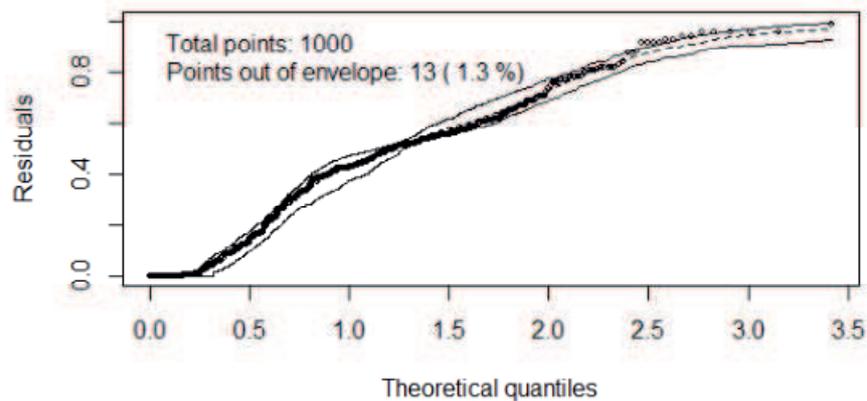
	McFadden	Likelihood	Cragg & Uhler's
Passo 1	0,1252971	0,2391788	0,2696045
Passo 2	0,09979649	0,19565186	0,22054047

A tabela 8 apresenta os valores do McFadden, *Likelihood* e Cragg & Uhler do modelo. Os dados evidenciam que as medidas *pseudo* R^2 aumentam à medida que previsoires são adicionados. A função do logaritmo de verossimilhança $-2LL$ reduz a cada passo (passo 1 modelo completo e passo 2 modelo reduzido à interação), ou seja, este teste objetiva verificar o poder de influência das variáveis independentes sobre a variável sintomas, de acordo com o critério de Cragg e Uhler o modelo logit estimado prediz bem sobre a variável resposta no estudo.

Na Figura 5 têm-se o gráficos *hnp* para o modelo, por meio deste, pode-se afirmar

que, o modelo apresenta um bom ajuste, pois o gráfico semi-normal com envelope simulado apresenta cerca de 1% dos valores fora das bandas de confiança.

Figura 5 – Modelo ajustado pela distribuição Multinomial.



De forma geral, em plantas onde foram aplicadas 50, 100 e 150 kg ha^{-1} de potássio, a incidência do tripses foi menor e significativamente diferentes das populações encontradas em plantas tratadas com a maior e menor dose do nutriente. A redução da incidência dos tripses, após 80 dias após emergência das plantas, pode estar relacionada a baixa emissão de brotos novos nas plantas tratadas com a maior e menor dose do nutriente. Os resultados afirmam que o potássio está envolvido no mecanismo de defesa das plantas a pragas e doenças, de modo que, altas concentrações de potássio promove a síntese e o acúmulo de compostos fenólicos, inibidores de insetos e fungos.

5 Conclusão

Neste estudo, observou-se que o efeito da adubação potássica sobre a planta de certa forma mesmo com ataque da praga resultou em uma diminuição de pontuações prateadas na planta. A partir das estimativas dos parâmetros do modelo logit multinomial, pode-se afirmar que houve efeito de interação entre o tratamento e as avaliações. Os valores de 50, 100 e 150 kg ha^{-1} de potássio, foram considerados os mais eficientes na redução das populações de adultos dos tripes. A incidência do tripes foi maior em plantas sem aplicação de potássio, o que proporcionou maior suscetibilidade ao ataque de pragas. O aumento das doses de potássio aplicadas causou a redução na população dos tripes nas plantas de amendoim. O que nos dá indícios de que as relações entre adubação de plantas e incidência de insetos podem indicar a quantidade em que um nutriente poderá favorecer ou impedir a ocorrência de insetos nas culturas.

Portanto verificou-se a importância da aplicação dos modelos lineares generalizados, pois se constitui de uma técnica bastante confiável para descrever o objeto em estudo. Nas análises realizadas o modelo que melhor se ajustou aos dados foi a regressão logística multinomial, conclui-se que presença de doses altas de potássio proporcionaram redução das populações de adultos e ninfas de *E. flavens* na cultura de amendoim, diminuindo consideravelmente os sintomas do tripes na planta.

Referências

- AKAIKE, H. A new look at the statistical model identification. *IEEE transactions on automatic control*, Ieee, v. 19, n. 6, p. 716–723, 1974. Citado na página 21.
- ATKINSON, A. Two graphical displays for outlying and influential observations in regression. *Biometrika*, Oxford University Press, v. 68, n. 1, p. 13–20, 1985. Citado na página 25.
- BARROS, F. U. *Matriz de covariâncias do estimador de máxima verossimilhança corrigido pelo viés em modelos lineares generalizados com parâmetro de dispersão desconhecido*. Tese (Doutorado) — Universidade Federal de Campina Grande, 2011. Citado na página 18.
- BATISTA, G.; GALLO, D.; CARVALHO, R. Determinação do período crítico de ataque do tripses do amendoim, *enneothrips flavens moulton*, 1941, em cultura “das águas”. *Anais da Sociedade Entomológica do Brasil*, v. 2, p. 45–53, 1973. Citado na página 10.
- BRITO, G. A. S.; NETO, A. A. Modelo de classificação de risco de crédito de empresas. *Revista Contabilidade & Finanças-USP*, Universidade de São Paulo, v. 19, n. 46, 2008. Citado na página 30.
- CALCAGNOLO, G.; LEITE, F.; GALLO, J. Efeitos da infestação do tripses nos folíolos do amendoim *enneothrips (enneothripiella) flavens moulton*, 1941, no desenvolvimento das plantas, na qualidade da produção de uma cultura “das águas”. *O Biológico*, Instituto Biológico., v. 40, n. 8, 1974. Citado na página 11.
- CASTRO, P. R. Análise de crescimento do amendoimzeiro (*arachis hypogaea* l.) com relação á infestação de pragas. *Anais da Escola Superior de Agricultura Luiz de Queiroz*, v. 31, p. 207–215, 1974. Citado na página 10.
- CORDEIRO, D. B. Ajuste de modelos lineares generalizados para dados positivos assimétricos. *Universidade Estadual da Paraíba*, p. 23–24, 2016. Citado na página 13.
- CORDEIRO, G. M.; NETO, E. d. A. Modelos paramétricos. *Pernambuco: UFRPE*, 2006. Citado 6 vezes nas páginas 11, 19, 20, 21, 22 e 23.
- GALLO D NAKANO, O. S. N. S. C. R. B. G. d. P. J. Z. R. A. S. V. J. M. L. o. Entomologia agrícola. piracicaba, fealq, 920p. *Macedo, Macedo, Campos, Novaretti e Ferraz*, v. 158, 2002. Citado na página 10.
- GREGORY W.C., K. A. M. o. Structure, variation, evolution, and classification in arachis. *Structure, variation, evolution, and classification in Arachis.*, Royal Botanic Gardens, Kew., p. 469–481, 1980. Citado na página 10.
- HABERMAN, S. J. Log-linear models for frequency tables with ordered classifications. *Biometrics*, JSTOR, p. 589–600, 1974. Citado na página 23.
- MACIEL, L. A. Uso de modelos lineares generalizados para avaliar táticas de manejo no controle do tripses do prateamento na cultura do amendoim. *Universidade Estadual da Paraíba*, 2016. Citado na página 17.

- MASCARENHAS HAA. PATRÍCIO, F. T. M. T. R. P. J. Ocorrência de fungos em sementes de soja produzidas sob calagem e adubação potássica residuais. *Scientia Agricola*, v. 52, n. 3, p. 426–430, 1995. Citado na página 11.
- MCCULLAGH, P.; NELDER, J. Generalised linear modelling. *Chapman and Hall: New York*, 1989. Citado 4 vezes nas páginas 13, 18, 22 e 23.
- MORAES, A. d. Efeito da infestação de *enneothrips flavens moulton* no desenvolvimento e produtividade de seis cultivares de amendoim, em condição de campo. 2005. 104f. *Instituto Agrônomo de Campinas*, 2005. Citado na página 24.
- MORAL, R. d. A. *Modelagem estatística e ecológica de relações tróficas em pragas e inimigos naturais*. Tese (Doutorado) — Universidade de São Paulo, 2013. Citado na página 25.
- MYERS RAYMOND H. MONTGOMERY, D. C. V. G. G. R. T. J. *Generalized linear models: with applications in engineering and the sciences*. [S.l.]: John Wiley & Sons, 2012. v. 791. Citado na página 21.
- NAKANO OCTAVIO SILVEIRA NETO, S. Z. R. *Entomologia econômica*. [S.l.]: USP-ESALQ, Departamento de Entomologia, 1981. Citado na página 10.
- NELDER, J. A.; WEDDERBURN, R. J. *Generalized linear models*. [S.l.]: Wiley Online Library, 1972. Citado 2 vezes nas páginas 12 e 18.
- NETER J KUTNER, K. M. N. Cj, wasserman, w. *Applied Linear Statistical Models*, v. 4, 1996. Citado na página 25.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2017. Disponível em: <<https://www.R-project.org/>>. Citado 2 vezes nas páginas 25 e 26.
- RENSI AA. CALCAGNOLO, G. O. D. [control of *enneothrips (enneothripiella) flavens moulton*, 1941, with organo-synthetic insecticides in a peanut crop during the rainy season [arachis hipogaea; thrips]].[portuguese]. *Biologico*, 1977. Citado na página 10.
- ROSSETO C.J., R. I. P. A. . I. T. Interação entre o tripes, a verrugose e variedades do amendoin—zeiro. *REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA PARA O PROGRESSO DA CIÊNCIA*, v. 20, n. 2, 1968. Citado na página 10.
- WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, JSTOR, v. 54, n. 3, p. 426–482, 1943. Citado na página 20.
- WEDDERBURN, R. W. Quasi-likelihood functions, generalized linear models, and the gauss—newton method. *Biometrika*, Oxford University Press, v. 61, n. 3, p. 439–447, 1972. Citado na página 12.

Apêndices

```
library(aod)
library(ggplot2)
library(Rcpp)
require(hnp)
require(reshape2)
dados=read.table("Potassio.txt",h=T)
head(dados)
attach(dados)
head(dados)
summary(dados)
sapply(dados, sd)

## Ajuste do modelo
z1 <- glm((SINTOMAS) ~BLOCO+TRAT+AVAL+AVAL:TRAT,
family = poisson(link="sqrt"),

data =dados)
z1
hnp(z1,print=T,verb=T)

library(nnet)
with(dados, table(AVAL, SINTOMAS,TRAT))
with(dados, do.call(rbind, tapply(SINTOMAS, TRAT,
function(x) c(M = mean(x), SD = sd(x)))))

with(dados, do.call(rbind, tapply(SINTOMAS, AVAL,
function(x) c(M = mean(x), SD = sd(x)))))

mod <- multinom(SINTOMAS ~ TRAT+AVAL, dados)

summary(mod)
predict(mod)
hnp(mod,print=T,verb=T)

#####

mod1 <- multinom(SINTOMAS ~ BLOCO+TRAT+AVAL+AVAL:TRAT, dados)
summary(mod1)
predict(mod1)
```

```
hnp(mod1, print=T, verb=T)
z <- summary(mod1)$coefficients/summary(mod1)$standard.errors
z

p <- (1 - pnorm(abs(z), 0, 1)) * 2
p

exp(coef(mod))

head(pp <- fitted(mod))

dses <- data.frame(TRAT = c(1,2,3,4,5), AVAL = mean(dados$AVAL))
predict(mod, newdata = dses, "probs")

dwrite <- data.frame(TRAT= rep(c(1,2,3,4,5), each = 10), AVAL =
rep(c(1,2,3,4,5,6,7,8,9,10),5))

pp.write <- cbind(dwrite, predict(mod, newdata = dwrite, type = "probs"
,se = TRUE))

by(pp.write[,3:6], pp.write$TRAT, colMeans)

lpp <- melt(pp.write, id.vars = c("TRAT", "AVAL"), value.name =
"probability")
head(lpp) # view first few rows

ggplot(lpp, aes(x = AVAL, y = probability, colour = TRAT)) + geom_line() +
facet_grid(variable ~ ., scales = "free")
```