



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Autorizada pelo Decreto Federal nº 77.496 de 27/04/76
Recredenciamento pelo Decreto nº 17.228 de 25/11/2016



PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

XXV SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS **SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2021**

MODELO DE REGRESSÃO POISSON NA ANÁLISE DOS DADOS DE UM EXPERIMENTO *IN VITRO* COM *COMANTHERA MUCUGENSIS*

Gerson Silva da Cunha¹; Mauricio Santana Lordelo²

1. Bolsista PROBIC/UEFS, Graduando em Agronomia, Universidade Estadual de Feira de Santana, gcunha.gavi@gmail.com
2. Orientador, Departamento de nome Ciências exatas, Universidade Estadual de Feira de Santana, mlordelo@uefs.br

PALAVRAS-CHAVE: modelos lineares generalizados; Poisson; indução *in vitro*

INTRODUÇÃO

O modelo clássico de regressão é a técnica mais adequada quando se deseja estudar o comportamento de uma variável dependente (variável resposta normalmente representada pela letra y) em relação a outras variáveis independentes (variáveis explicativas normalmente representadas pela letra x), que são responsáveis pela variabilidade da variável resposta.

A estrutura, que pode ser expressa em forma matricial, é formada por um vetor de variáveis independentes, de variáveis dependentes e de parâmetros a serem estimados. O processo de estimação dos parâmetros implica em considerar um vetor de erros que se assume ter média zero, mesma variabilidade e independência, ou pelo menos não correlacionados. No entanto, nem sempre estas suposições são satisfeitas.

A ausência das suposições básicas implica na obtenção de resultados e interpretações muitas vezes inadequadas para o pesquisador. Dessa forma, uma alternativa é o uso dos modelos lineares generalizados (MLG's). Estes modelos são uma extensão dos modelos lineares clássicos em que uma série de técnicas comumente estudadas separadamente foram reunidas em uma única teoria (Nelder e Wedderburn, 1972). Permitem modelar dados em situações onde a distribuição de probabilidade da variável resposta pertence a uma família de variáveis, definida como família exponencial, da qual fazem parte a distribuição Normal, Binomial, Poisson, Gama, dentre outras.

A distribuição Poisson é um modelo probabilístico largamente utilizado em experimentos de contagens de plantas, colônias de bactérias, itens, objetos, dados em intervalo de tempo, área, volume ou comprimento (Andrade & Ogliari, 2010). Quando se deseja fazer um estudo sobre a distribuição de uma espécie vegetal em um habitat e supondo que a distribuição dessa espécie neste habitat é considerada aleatória (casual) o modelo de Poisson é apropriado. Muito comum em estudos de seguimento, onde as unidades experimentais são classificadas segundo os níveis de categorias, tais como tipo de tratamento, tipo de cultura, tipo de genótipo, etc., e são acompanhadas por um

período fixo pré-estabelecido ou até a ocorrência de um determinado evento. Dados dessa natureza são caracterizados por não atenderem aos pressupostos básicos do modelo clássico de regressão, justificando o uso dos MLG's.

MATERIAL E MÉTODOS

Um experimento foi desenvolvido para avaliar o efeito de diferentes concentrações de BAP na indução *in vitro* de brotos da espécie *Comanthera mucugensis*. Plântulas desta espécie, germinadas *in vitro*, foram utilizadas como explante e inoculadas em tubos de ensaio contendo 10 ml de meio de cultura MS com metade das concentrações salinas acrescido de 15 gL⁻¹ de sacarose, 7 gL⁻¹ de ágar e diferentes concentrações do regulador vegetal 6-benzilaminopurina – BAP (0,00; 1,10; 1,32; 1,54; 1,76). O delineamento foi inteiramente casualizado, com cinco tratamentos, cada um composto por 18 unidades experimentais. Cada unidade representa um tubo contendo uma plântula. Após 60 dias foram registrados o número de brotos por explante (NB). Essa variável é classificada como de contagem e com características próximas de uma distribuição Poisson. Dessa forma, foram ajustados diferentes modelos para dados dessa natureza e comparados por meio da análise dos resíduos. Optou-se pelo uso dos gráficos de probabilidade Normal com um envelope simulado desenvolvido por Moral et al. (2017). Esses gráficos são um indicador útil de adequação que pode ser usado com qualquer modelo linear generalizado e extensões. Foi usado o *software* livre R para gerar os resultados (R Core Team, 2020)

RESULTADOS E DISCUSSÃO

Uma análise inicial, mostrou que o número médio aproximado de brotos e a variância foram, respectivamente, 14,8 e 618,4. Através da Figura 1, é possível observar que os dados apresentam uma grande assimetria e um grande número de zeros.

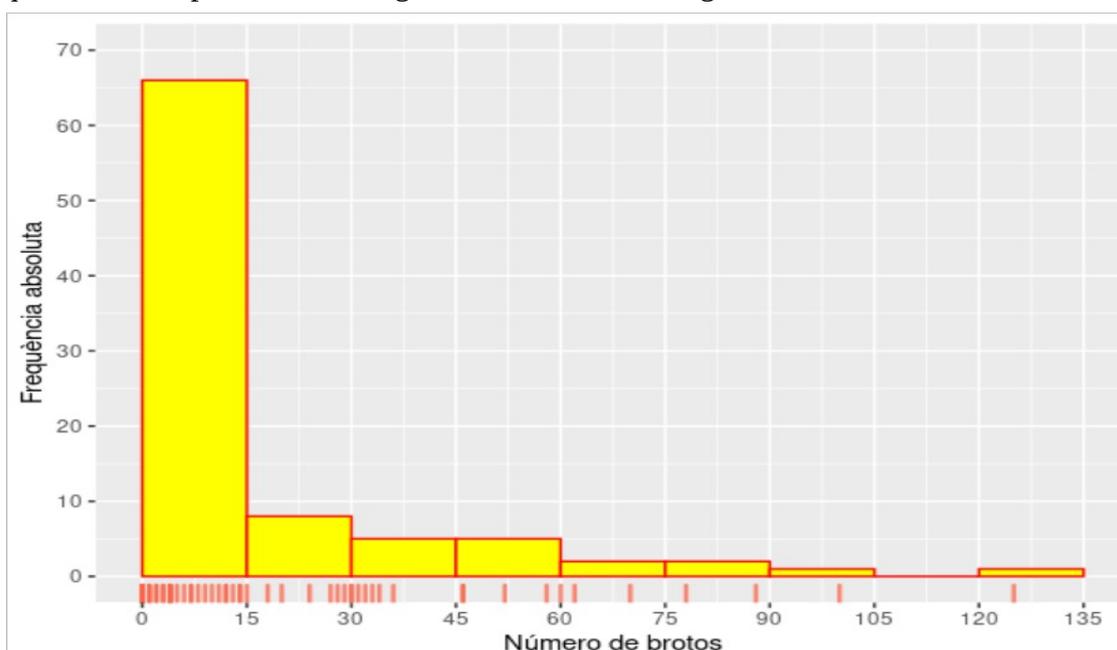


Figura 1: Distribuição do número de brotos em intervalos.

Os dados apresentaram uma grande variabilidade, tanto os dados de um modo geral, quanto dentro de cada tratamento, como mostra a Tabela 1.

Tabela 1: Número de brotos por concentração administrada.

| Medida | Concentração de BAP (μM) | | | | |
|-----------|---------------------------------------|---------|----------|---------|---------|
| | 0,00 | 1,10 | 1,32 | 1,54 | 1,76 |
| Média | 0,889 | 14,555 | 24,778 | 17,833 | 16,000 |
| Variância | 6,810 | 515,908 | 1461,124 | 463,676 | 468,588 |

Para efeito de comparação, primeiro ajustou-se o modelo clássico aos dados. Como era esperado, os resultados não atenderam aos pressupostos de normalidade dos resíduos e homogeneidade de variâncias pelos testes de *Shapiro-Wilk* e *Levene*.

Em seguida, procedeu-se o ajuste do modelo de Poisson utilizando a função de ligação logarítmico. Pelo fato da variância ser muito maior do que a média, indicando uma superdispersão, o ajuste deste modelo também não apresentou resultados satisfatórios. Devido a superdispersão, ocasionado pelo elevado valor da variância em comparação com a média, foi necessário recorrer aos modelos de Poisson estendidos. Estes foram originalmente desenvolvida por Faddy (1997), onde foi descrita a construção de distribuições de probabilidade discretas com propriedades de dispersão muito gerais. Assim, mais duas opções de modelagem para dados com essa característica foram usados: Quasi-Poisson e Binomial Negativa.

Na Figura 2, tem-se o comportamento dos resíduos para os quatro modelos.

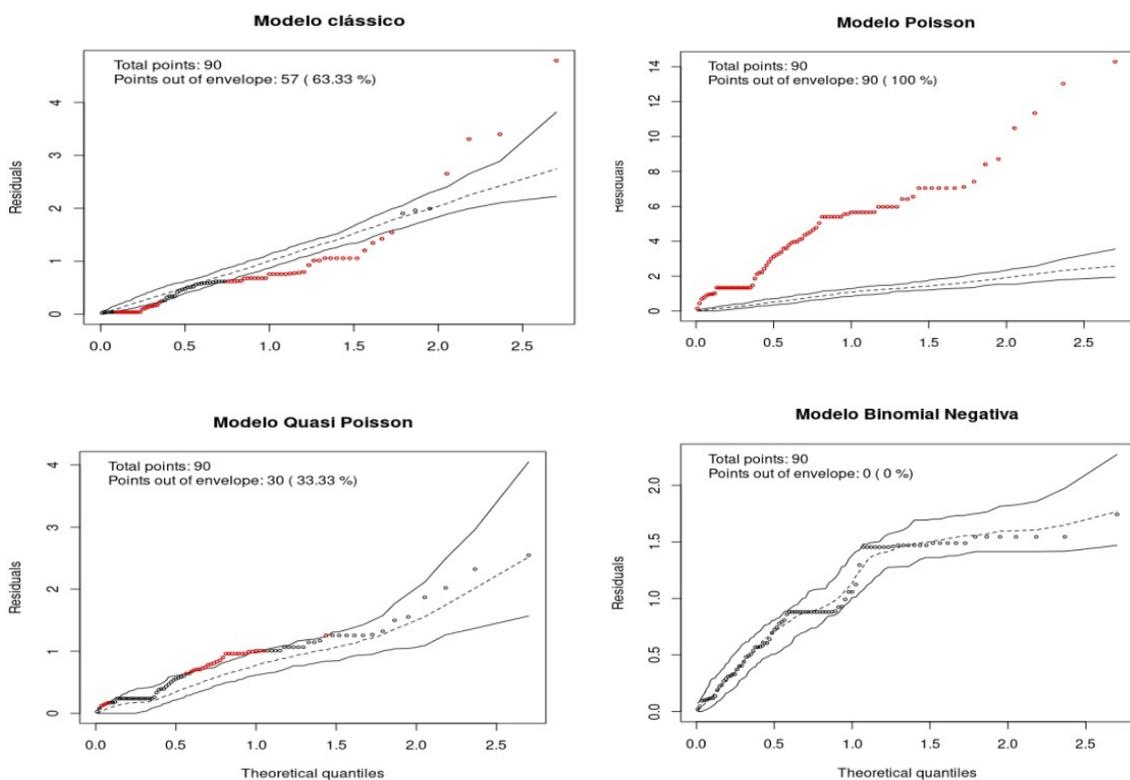


Figura 2: Distribuição dos resíduos obtidos por cada um dos modelos ajustados.

A avaliação de cada um dos modelos, leva em consideração a quantidade de pontos que se distribuem dentro do chamado envelope simulado. Quanto mais pontos dentro desse envelope, melhor o ajuste do modelo em questão. Observa-se então, que o modelo usando a distribuição Poisson teve o pior resultado uma vez que todos os pontos ficaram fora do envelope. Já o modelo usando a distribuição Binomial Negativa teve assim o melhor ajuste. Recomenda-se o uso deste modelo para realização das inferências (testes de hipóteses e intervalos de confiança) a fim de verificar as possíveis diferenças estatísticas entre as concentrações testadas do regulador BAP sobre o número de brotos por explante.

CONSIDERAÇÕES FINAIS

Foi possível observar a importância de se fazer uma investigação bastante cuidadosa sobre a natureza dos dados coletados, para uma posterior avaliação das possíveis diferenças estatisticamente significativas entre tratamentos. Os resultados mostraram que é fundamental uma análise descritiva prévia, para conhecer melhor os dados e suas características e, assim, tomar as devidas decisões. Através do gráfico com envelope simulado, foi possível verificar a diferença expressiva que existe entre a qualidade de ajuste dos quatro modelos testados. O uso de modelos cujo o ajuste não foi adequado, tende a comprometer a qualidade e robustez das inferências, levando a conclusões que não condizem com a verdadeira realidade.

REFERÊNCIAS

- ANDRADE, D.F., OGLIARI, P.J. Estatística para ciências agrárias e biológicas - com noções de experimentação. 3ª ed. UFSC editora. 2010.
- FADDY, M.J (1997). “Extended Poisson Process Modelling and Analysis of Count Data”. *Biometrical Journal*, 39(4), 431–440. doi:10.1002/bimj.4710390405.
- MORAL, R.A; HINDE, J; DEMÉTRIO, C.G.B. (2017). “Half-Normal Plots and Overdispersed Models in R: The hnp Package”. *Journal of Statistical Software*, 81(10), 1–23. doi: 10.18637/jss.v081.i10.
- NELDER, J.A.; WEDDERBURN, R.W. Generalized linear models.. *Journal of the Royal Statistical Society*, London, v. 135, p. 370-384, 1972.
- R CORE TEAM (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.