



Análise de dados sobre de germinação de sementes via GAMLSS

Edilenia Queiroz Pereira¹, Oilson Alberto Gonzatto Junior ¹, Marcos Jardel Henriques¹
Vanderly Janeiro ² Terezinha Aparecida Guedes ² Jose Walter Pedroza Carneiro ³

¹ Discente Pós-graduação em Bioestatística – UEM

² Docente Departamento de Estatística - UEM

³ Docente Departamento de Agronomia - UEM

RESUMO

Com base nos dados experimentais, obtidos por pesquisadores da Fazenda Experimental de Iguatemi, estudou-se o comportamento da germinação da planta *Stevia rebaudiana* (Bert.) Berton (estévia). O modelo de regressão não-linear misto com a distribuição Binomial mostrou que é adequada para explicar a germinação das sementes da planta supracitada, com relação aos fatores de influência ou variáveis explicativas, a saber luz, posição, lote e germinador. A estimação e validação do modelo foi realizada por meio do pacote “gamlss” do ambiente estatístico R. Com base no modelo ajustado foi possível identificar que não existe diferença de germinação das sementes entre os quatros germinadores utilizados. Por outro lado, tanto o tipo de luz aplicada como a posição em que se aloca essas sementes dentro dos germinadores mostraram estar influenciando no tipo de germinação das sementes da planta *estévia*.

Palavras chave: Stevia, Regressão logística mista, GAMLSS.

1 INTRODUÇÃO

A busca pelo entendimento da variável resposta com relação as variáveis explicativas, faz com que os modelos de regressões sejam de grande importância em muitos problemas de diversas áreas do conhecimento como na agrônômica, médica, biológica, industrial, química entre outras. Uma aplicação factível desses modelos está no estudo da germinação de sementes com relação aos possíveis fatores que podem influenciar esse processo. Desta forma, a baixa germinação e conseqüentemente a dificuldade de produção em grande escala da planta conhecida por *estévia*, motivou pesquisadores da Universidade

Estadual de Maringá (UEM) a estudar e realizar experimentos sobre condições controladas para tentar entender as melhores condições que favorecem a germinação deste tipo de semente.

Assim, o objetivo deste trabalho é apresentar uma modelagem estatística adequada referente aos dados hierárquicos sobre germinação de sementes de *estévia* por meio da distribuição Binomial. O ajuste deste modelo foi realizado por meio do ambiente estatístico R (R Core Team, 2016), por meio do pacote “*gamlss*”.

2 Modelo de regressão Binomial misto

Os modelos de regressões lineares generalizados mistos são modelos matemáticos que buscam descrever o comportamento de uma variável resposta, com relação a uma ou mais variáveis explicativas, as quais podem envolver tanto efeitos fixos como aleatórios. Associada a essa variável resposta tem uma distribuição de probabilidade que busca relacionar os valores observados com as suas respectivas probabilidade de ocorrência.

Sendo Y_1, Y_2, \dots, Y_n uma amostra aleatória independente da variável aleatória Y , tal que $y_i \sim \text{Bin}(n_i, \pi_i)$. O modelo de regressão linear generalizado misto com efeitos mistos na forma matricial dada por:

$$\log\left(\frac{\pi}{1-\pi}\right) = \eta = \underbrace{\mathbf{X}\boldsymbol{\beta}}_{\text{Fixo}} + \underbrace{\mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}}_{\text{Aleatório}}, \quad (1)$$

em que;

$g(\boldsymbol{\pi})$ é a função de ligação que relaciona a parâmetro $\boldsymbol{\pi}$ da distribuição $\text{Bin}(\mathbf{n}, \boldsymbol{\pi})$ com o seu preditor linear $\boldsymbol{\eta}$;

\mathbf{X} é a matriz $n \times (1+p)$ de valores conhecidos referente ao intercepto e as p covariáveis de efeito fixo;

\mathbf{Z} é a matriz $n \times q$ de valores conhecidos de q covariáveis de efeito aleatório;

$\boldsymbol{\beta}$ é o vetor $(1+p) \times 1$ de parâmetros estimados referente aos efeitos fixos;

\mathbf{b} é o vetor $q \times 1$ de parâmetros estimados referente aos efeitos aleatórios, em que assume-se com distribuição multivariada $N(\mathbf{0}, \mathbf{D})$;

$\boldsymbol{\varepsilon}$ é um vetor $n \times 1$ de erros aleatórios associado ao modelo com distribuição multivariada $N(\mathbf{0}, \mathbf{R})$.

Sabendo que, $g(\pi_i)$ representa função de ligação adequada, utilizada para relacionar o parâmetro π_i ao seu preditor linear η_i , tem-se que a média e variância da variável aleatória y_i após a regressão podem ser calculadas por:

$$E(y_i) = ng^{-1}(\eta_i) \text{ e } \text{Var}(y_i) = ng^{-1}(\eta_i)(1 - g^{-1}(\eta_i))$$

Assim, as estimativas intervalares para $E(y_i)$ e $\text{Var}(y_i)$, podem ser obtidas por meio do Método Delta. Mais detalhes sobre modelos mistos não lineares estão descritos em Pinheiro e Bates (2006) e Fitzmaurice, Laird e Ware (2012).

3 Descrição do banco de dados

Com o intuito de identificar as melhores condições para a germinação da semente de *estévia*, sob condições controladas, pesquisadores realizaram em 2013 um experimento

na Fazenda Experimental de Iguatemi (FEI), que está localizada no Distrito de Iguatemi, Município de Maringá, Paraná, Brasil. Neste experimento foram utilizados 4 germinadores, 2 tipos de luz, 9.600 sementes de *estévia* pertencentes a 4 lotes de sementes distintos, 96 bandejas de plástico branco e fibra de algodão farmacêutica. Cada uma das 96 bandejas foram preenchidas com volume de 4 cm³ com fibras de algodão farmacêutica e colocada para germinar 100 sementes de *estévia*. Em cada um dos germinadores foram colocadas 24 bandejas brancas, as quais foram distribuídas aleatoriamente, 8 na parte superior, 8 na parte média e 8 na parte inferior. Cada germinador só tinha sementes pertencentes a dois lotes distintos, desta forma das 8 bandejas 4 eram referentes a replicação de um lote de sementes e as outras 4 a replicação de outro lote, assim . O experimento teve duração de 10 dias, sendo que no décimo dia foi contada e registrada, em cada uma das bandejas, a quantidade de embriões vegetais normais, anormais e sementes mortas.

4 Resultados e Discussões

4.1 Modelo proposto

Por meio do modelo de regressão Binomial ajustou-se cinco modelos distintos com a finalidade de identificar qual descreve melhor a relação existente entre a germinação das sementes de *estévia* sob os fatores de influência: como posição no germinador, luz, lote e germinação. A seleção do modelo foi baseada nas medidas referentes ao Critério de Informação de Akaike (AIC), a Deviance Global e pelo Critério de Informação Bayesiana (BIC), sendo que os valores destas medidas para cada um dos modelos propostos estão sendo apresentado na Tabela 1.

Tabela 1: Medidas para comparação e seleção dos modelos ajustados, com função de ligação logística.

Modelos	AIC	D. Global	BIC
$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3$	1401,48	1395,48	1409,17
$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 luz$	1364,70	1356,70	1374,96
$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 luz + b_1 Lote$	722,55	706,62	742,96
$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 luz + b_1 germinador$	1251,23	1236,41	1270,22
$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 luz + b_1 lote + b_2 germinador$	724,55	706,62	747,53

Conforme os resultados da Tabela 1 pode-se perceber que o modelo $\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 luz + b_1 Lote$ é o que apresentou o menor valor de AIC, a menor Deviance Global e o menor BIC entre todos os modelos ajustados, sendo então esse o modelo escolhido. Em seguida, realizou-se a análise de resíduos e obteve-se resultados satisfatórios, validando assim o modelo adotado.

A Tabela 2 apresenta todas as estimativas do modelo selecionado com os seus respectivos desvios padrão e testes de significância.

Tabela 2: Estimativas, erros padrões e teste t para os parâmetros do modelo de regressão Binomial adotado.

Variável relacionada	Parâmetro	Estimativa	Erro Padrão	Valor t	Pr(> t)
-	β_0	0,49	0,04	11,23	0,00
Posição 2	β_1	-0,01	0,05	-0,16	0,87
Posição 3	β_2	-0,27	0,05	-4,96	0,00
Luz 2	β_3	-0,23	0,04	-5,24	0,00
Lote	σ_b^2	0,93	0,99	-	-

Os resultados contidos na Tabela 3, mostram que as sementes de *estévia* quando são colocadas na posição 2 tem a mesma chance de germinar normal quando comparado com a posição 1. A chance de uma semente colocada na posição 3 germinar normal é equivalente a 0,77 vezes a chance de uma planta germinar normal na posição 1. A luz também está influenciando no tipo de germinação, pois foi identificado que a chance de germinar normal para sementes sobre o efeito da luz 2 equivale a 0,79 vezes a chance de germinar normal sobre efeito da luz 1.

Tabela 3: Razões de chances e o seus respectivos intervalos de confiança, referente ao modelo de regressão Binomial.

	OR	IC(OR,95%)
Posicao2	0,99	(0,8918; 1,1019)
Posicao3	0,77	(0,6894; 0,8511)
Luz2	0,79	(0,6894; 0,8511)

5 Conclusão

De acordo com os resultados obtidos foi possível observar que o modelo de regressão linear generalizados com distribuição Binomial foi adequado para explicar o comportamento da germinação das sementes da planta *estévia* com relação aos fatores de influência. Em particular o tipo de luz e a posição da bandeja se mostram significativos na proporção de sementes germinada.

Referências

FITZMAURICE, G. M.; LAIRD, N. M.; WARE, J. H. *Applied longitudinal analysis*. New Jersey: John Wiley & Sons, 2012.

PINHEIRO, J.; BATES, D. *Mixed-effects models in S and S-PLUS*. New York: Springer Science & Business Media, 2006.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2016. Disponível em: <<https://www.R-project.org/>>.