



Edilenia Queiroz Pereira

**Modelos de regressão com efeitos mistos  
aplicados a dados de germinação de sementes**

Maringá - PR  
Março de 2017

Edilenia Queiroz Pereira

## **Modelos de regressão com efeitos mistos aplicados a dados de germinação de sementes**

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito para obtenção do título de Mestre em Bioestatística.

Universidade Estadual de Maringá – UEM

Departamento de Estatística

Programa de Pós-Graduação em Bioestatística

Orientador: Prof. Dr. Vanderly Janeiro

Coorientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Terezinha Aparecida Guedes

Maringá - PR

Março de 2017

Dados Internacionais de Catalogação na Publicação (CIP)  
(Biblioteca Central - UEM, Maringá, PR, Brasil)

P436m Pereira, Edilenia Queiroz  
Modelos de regressão com efeitos mistos aplicados  
a dados de germinação de sementes / Edilenia Queiroz  
Pereira. -- Maringá, 2017.  
67 f. : il. color., figs., tabs.

Orientador: Prof. Dr. Vandelry Janeiro.  
Coorientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Terezinha Aparecida  
Guedes.

Dissertação (mestrado) - Universidade Estadual de  
Maringá, Departamento de Estatística, Programa de  
Pós-Graduação em Biostatística, 2017.

1. Stevia rebaudiana (Bert) Bertoni - Modelo  
GAMLSS. 2. Stevia rebaudiana (Bert) Bertoni - Razão  
de Chance. 3. Stevia rebaudiana (Bert) Bertoni -  
Germinação de sementes - Modelos. 4. Bioestatística  
- Stevia rebaudiana (Bert) Bertoni. I. Janeiro,  
Vandelry, orient. II. Guedes, Terezinha Aparecida,  
coorient. III. Universidade Estadual de Maringá.  
Departamento de Estatística. Programa de Pós-  
Graduação em Biostatística. IV. Título.

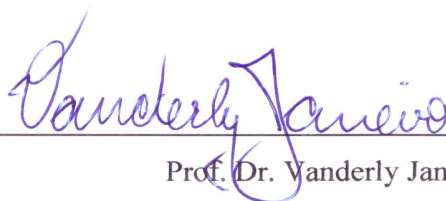
CDD 21.ed. 631.523

**EDILENIA QUEIROZ PEREIRA**

**Modelos de regressão com efeitos mistos aplicados a dados de  
germinação de sementes**

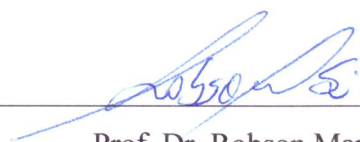
Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

**BANCA EXAMINADORA**



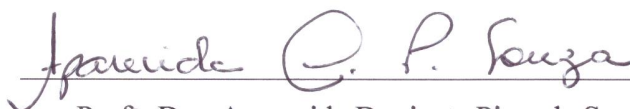
Prof. Dr. Vanderly Janeiro

Universidade Estadual de Maringá – PBE/UEM



Prof. Dr. Robson Marcelo Rossi

Universidade Estadual de Maringá – PBE/UEM



Profª. Dra. Aparecida Donizete Pires de Souza

Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP

Maringá, 20 de março de 2017.

*Ao nosso bebê Lenz, bem precioso que pintou de alegria, de cores e companhia os meus dias nessa reta final do mestrado, dedico.*

---

# AGRADECIMENTOS

---

A Deus e a Nossa Senhora Aparecida, em quem sempre depus minha fé e onde sempre pude encontrar paz, segurança e força. Estes últimos anos me ajudou a fortalecer ainda mais minha fé, pois em muitos momentos tive que fazer a fé pensar e a razão crer.

Ao meu esposo Antônio Oliveira por ter incentivado, apoiado, sonhado e acreditado em mim para a realização de mais uma conquista. Aos meus pais, Ildete Queiroz e Manoel Lacerda pelo dom da vida, presença constante, mesmo distante, em todos os passos de minha trajetória, apoio incondicional em todas as minhas decisões, amor e orações. Ao meu irmão Enio Queiroz e minha cunhada Andreia Brandão por sempre estarem ao meu lado para o que eu precisar. Tenho certeza que sem vocês ao meu lado eu não seria nada e nem teria chegado até aqui, sou eternamente grata. Amo muito vocês!

Ao meu orientador Prof. Dr. Vandely Janeiro e a minha Co-orientadora Prof<sup>a</sup>. Dr<sup>a</sup>. Terezinha Aparecida Guedes pelo acolhimento, paciência, ensinamentos, confiança, presença, orientando-me nas minhas limitações e dúvidas. A orientação de vocês foram de extrema importância na construção do meu aprendizado, para o meu crescimento profissional e para a realização deste trabalho. E também ao Prof. Dr. José Walter P. Carneiro pela parceria disponibilizando os dados, contribuindo com o seu inglês e com as valiosas discussões.

A todos os meus colegas e amigos da Estatística pelos momentos compartilhados juntos durante esse tempo em que estive na UEM, em especial a Hellen Geremias, Juliana Nascimento, Márcia Lorena e Oilson Granzoto por nunca terem dito não a um pedido de ajuda.

A todos os professores do Departamento de Estatística e do Programa de Pós-graduação em Bioestatística pelo valioso conhecimento transmitido.

Ao Programa de Pós-graduação em Bioestatística pelo acolhimento, especialmente a Prof<sup>a</sup>. Dr<sup>a</sup>. Eniuce Menezes por ser sempre muito prestativa.

A todos da minha família pelo amor, torcida e apoio. Especialmente aqueles que de alguma forma me ajudou nos momentos que precisei.

A minha sogra Alvina e a todos da família Oliveira pela torcida e apoio.

E por fim, a CAPES, pelo suporte financeiro.

*“O Correr da vida embrulha tudo. A vida é assim: esquenta e esfria, aperta e daí afrouxa, sossega e depois desinquieta. O que ela quer da gente é coragem...”*

*(Guimarães Rosa)*

# RESUMO

A *Stevia rebaudiana* é uma planta que tem grande potencial quando se trata da produção de adoçantes naturais. Esse fato produz um grande interesse econômico e científico com relação as suas características e propriedades terapêuticas, as quais estão presentes em grande maioria nas suas folhas. Interessados em entender o comportamento da germinação das sementes de *Stevia rebaudiana*, sobre condições controlada, fez com que pesquisadores da Fazenda Experimental de Iguatemi realizasse um experimento com 9.600 sementes retiradas de quatro lotes distintos e submetidas sobre efeitos de luz e posição de locação dentro do germinador utilizado. A fim de analisar esses dados coletados empregou-se os modelos de regressão lineares generalizados mistos com distribuição Binomial, Beta-binomial com abordagem frequentista e bayesiana e a distribuição Multinomial com abordagem frequentista. A estimação e validação dos modelos ajustados foram realizadas através dos pacotes “gamlss” frequentista e “INLA” bayesiano do ambiente estatístico R. Empregou-se a função de ligação *logit* em todos os modelos ajustados e a escolha do modelo final deu-se por meio das medidas AIC, BIC e Deviance Global do ponto de vista frequentista e DIC, WAIC e LPML do ponto de vista bayesiano. Pelos critérios de seleção de modelos percebeu-se que os modelos Binomial e Beta-binomial não apresentaram diferença com relação a qualidade dos resultados obtidos, sendo assim equivalentes, tanto pelo ponto de vista frequentista quanto bayesiano. A metodologia empregada mostrou-se adequada na explicação da germinação, com relação aos fatores de influência ou variáveis explicativas, a saber luz, posição, lote e germinador. Com base nos modelos ajustados foi possível identificar que não existe diferença de germinação das sementes entre os quatro germinadores utilizados. Por outro lado, tanto o tipo de luz aplicada como a posição em que se coloca essas sementes dentro dos germinadores mostraram influenciar os resultados de germinação das sementes.

**Palavras-chaves:** Modelo GAMLSS, Razão de Chance, *Stevia rebaudiana*.



# ABSTRACT

*Stevia rebaudiana* is a plant that has great potential when it comes to the production of natural sweeteners, this fact produces a great economic and scientific interest in relation to its characteristics and therapeutic properties, which are present in great majority in its sheets. Interested in understanding the germination behavior of *Stevia rebaudiana* seeds under controlled conditions, researchers from the Experimental Farm of Iguatemi conducted an experiment with 9,600 seeds taken from four different lots and submitted to light and position effects Within the germinator us. In order to analyze these data, we used the generalized linear mixed regression models with Binomial distribution, Beta-binomial with frequentist approach and Bayesian approach and the Multinomial distribution with frequentist approach. The estimation and validation of the adjusted models were performed by the frequentist package "gamlss" and by Bayesian package "INLA" of the statistical environment R. The logit link function was used in all fitted models and the choice of the Final model was given by means of AIC, BIC and Deviance Global measurements from the frequentist point of view and DIC, WAIC and LPML from the Bayesian point of view. By the criteria of selection of models it was noticed that the Binomial and Beta-binomial models did not present difference in relation to the quality of the obtained results, being thus equivalent, both by frequentist point of view as Bayesian. The methodology used was adequate in explaining the germination, in relation to the factors of influence or explanatory variables, namely light, position, lot and germinator. Based on the adjusted models it was possible to identify that there is no difference in seed germination among the four germinators used. On the other hand, both the type of light applied and the position in which these seeds are placed inside the germinators have been influencing the germination results of the seeds.

**Key-words:** Model GAMLSS, Odds ratio, *Stevia rebaudiana*.

---

## LISTA DE ILUSTRAÇÕES

---

Figura 1 – Parcela de campo cultivada (a) e parcela em vaso cultivado (b) com <i>Stevia rebaudiana</i> (Bert.) Bertoni. . . . .	18
Figura 2 – Delineamento do experimento realizado com as sementes de <i>Stevia rebaudiana</i> . . . . .	19
Figura 3 – Características morfo-anatômicas das plântulas normais (a), plântulas anormais (b) e sementes não germinadas (c). . . . .	20
Figura 4 – Função distribuição de probabilidade da Binomial para valores de $\pi$ iguais a 0,2; 0,5; 0,8 e $n = 100$ . . . . .	21
Figura 5 – Função distribuição de probabilidade da Beta-binomial com valores de $\pi$ iguais a 0,1; 0,3; 0,5; 0,7 e 0,9 e valores de $\sigma$ iguais a 0,01; 0,10; 0,50 e 1,00. . . . .	23
Figura 6 – Função distribuição de probabilidade da Multinomial para $\pi_1$ e $\pi_2$ iguais a 0,1; 0,3 e 0,4, sendo que a dimensão da bola representa a função de densidade nos pontos, para $n = 5$ . . . . .	25
Figura 7 – Representação de como estão alocados os modelos de regressão. . . . .	28
Figura 8 – <i>Worm plot</i> de um modelo com ajuste adequado. . . . .	33
Figura 9 – Proporção do tipo de plântulas normais, anormais e sementes não germinadas observadas no experimento. . . . .	39
Figura 10 – Comportamento do tipo de plântulas normais (N), anormais (AN) e sementes não germinadas (NG) dentro do experimento levando em consideração os fatores como germinador, luz, posição e lote. . . . .	40
Figura 11 – Categorização utilizada do tipo de plântulas produzidas para cada uma das distribuições utilizadas. . . . .	41
Figura 12 – Gráficos para diagnóstico do ajuste do modelo considerando a distribuição Binomial para os dados. . . . .	43
Figura 13 – <i>Worm plot</i> do resíduo quantílico aleatorizado referente a regressão Binomial. . . . .	44
Figura 14 – (a) Gráfico dos valores de PIT para o modelo Binomial. (b) Histograma dos valores de PIT para o modelo Binomial. . . . .	45

Figura 15 – Densidades marginais <i>a posteriori</i> para $\beta_0, \beta_1, \beta_2$ e $\beta_3$ . . . . .	45
Figura 16 – Gráficos para diagnóstico do ajuste do modelo considerando a distribuição Beta-binomial para os dados. . . . .	47
Figura 17 – <i>Worm plot</i> do ajuste com a distribuição Beta-binomial. . . . .	48
Figura 18 – (a) Gráfico dos valores de PIT para o modelo Beta-binomial. (b) Histograma dos valores de PIT para o modelo Beta-binomial. . . . .	49
Figura 19 – Densidades marginais <i>a posteriori</i> para $\beta_0, \beta_1, \beta_2$ e $\beta_3$ . . . . .	49
Figura 20 – Gráficos para diagnóstico do ajuste do modelo considerando a distribuição Multinomial para os dados. . . . .	51
Figura 21 – <i>Worm plot</i> do ajuste com a distribuição Multinomial. . . . .	52

---

## LISTA DE TABELAS

---

Tabela 1 – Descrição das variáveis explicativas utilizadas no experimento. . . . .	19
Tabela 2 – Funções de ligação. . . . .	29
Tabela 3 – Descrição dos formatos que o <i>Worm plot</i> pode assumir. . . . .	33
Tabela 4 – Modelos de regressão a serem ajustados considerando as distribuições Binomial, Beta-binomial e Multinomial. . . . .	41
Tabela 5 – Critérios frequentistas e bayesianos para comparação e seleção dos modelos ajustados com a distribuição Binomial. . . . .	42
Tabela 6 – Estimativas pontuais e intervalares do ponto de vista frequentista e bayesiano para os coeficientes de regressão utilizados para ajustar o parâmetro $\pi$ da distribuição $BI(n, \pi)$ . . . . .	42
Tabela 7 – Medidas descritivas do resíduo quantílico aleatorizado. . . . .	44
Tabela 8 – Medidas referente a abordagem frequentista e bayesiana na comparação e seleção dos seis modelos ajustados com a distribuição Beta-binomial. . . . .	46
Tabela 9 – Estimativas pontuais e intervalares do ponto de vista frequentista e bayesiano para os coeficientes de regressão utilizados para ajustar o parâmetro $\pi$ da distribuição $BB(\pi, \sigma)$ . . . . .	46
Tabela 10 – Estimativas pontuais e intervalares do ponto de vista frequentista e bayesiano para os coeficientes de regressão utilizados para ajustar o parâmetro $\sigma$ da distribuição $BB(\pi, \sigma)$ . . . . .	47
Tabela 11 – Resumo do resíduo quantílico aleatorizado. . . . .	48
Tabela 12 – Medidas para comparação e seleção dos modelos ajustados com a distribuição Multinomial. . . . .	50
Tabela 13 – Estimativas pontuais e intervalares do modelo de regressão Multinomial referente a proporção de plântulas normais com relação a sementes não germinadas e da proporção de plântulas anormais com relação a sementes não germinadas. . . . .	50
Tabela 14 – Resumo do resíduo quantílico aleatorizado. . . . .	51

Tabela 15 – Medidas de seleção de modelo referente ao modelo $M_3$ ajustado com as distribuições Binomial, Beta-binomial e Multinomial. . . . .	52
Tabela 16 – Razões de chances (OR) estimadas, intervalos de confiança e e intervalo de credibilidade, referente ao modelo de regressão Binomial e Beta-binomial.	53
Tabela 17 – Razões de chances estimadas (OR) e os seus respectivos intervalos de confiança do modelo de regressão Multinomial ajustado com relação a proporção de sementes com plântulas normais em relação as sementes não germinadas e a proporção de plântulas anormais com relação as sementes não germinadas. . . . .	54

---

# SUMÁRIO

---

<b>1</b>	<b>Introdução</b>	<b>14</b>
<b>2</b>	<b>Materiais e métodos</b>	<b>18</b>
2.1	Materiais	18
2.2	Métodos	20
2.2.1	Modelos de regressão linear generalizado misto	20
2.2.1.1	Modelo de regressão Binomial misto	20
2.2.1.2	Modelo de regressão Beta-binomial misto	22
2.2.1.3	Modelo de regressão Multinomial misto	24
2.2.1.4	Método Delta	27
2.2.2	Abordagem Frequentista	28
2.2.2.1	Modelo aditivo generalizado para locação, escala e forma	28
2.2.2.1.1	Método de estimação	29
2.2.2.1.2	Seleção do modelo	31
2.2.2.1.3	Diagnóstico	31
2.2.3	Abordagem Bayesiana	33
2.2.3.1	Teorema de Bayes	34
2.2.3.2	Distribuição <i>a priori</i> e <i>a posteriori</i>	34
2.2.3.3	Aproximação de Laplace Aninhada Integrada (INLA)	35
2.2.3.3.1	Seleção e validação de modelo	36
<b>3</b>	<b>Resultados e Discussões</b>	<b>39</b>
3.1	Análise exploratória	39
3.2	Resultados	40
3.2.1	Regressão Binomial	41
3.2.2	Regressão Beta-binomial	46
3.2.3	Regressão Multinomial	50
3.2.4	Discussões	52
<b>4</b>	<b>Considerações finais</b>	<b>56</b>
	Referências	57
	ANEXO A Códigos Inferência Frequentista	60
	ANEXO B Códigos Inferência Bayesiana	67

---

## CAPÍTULO 1

---

# INTRODUÇÃO

---

A germinação de semente é um fenômeno biológico que permite o desenvolvimento das estruturas essenciais do embrião, manifestando assim a sua capacidade de dar origem a novas plântulas normais, sob condições ambientais favoráveis (NASSIF; VIEIRA; FERNADES, 1998). Porém, nem sempre essa germinação permite uma geração de plântula normal, isso pode ser motivado por alguns fatores externos como a luz, a temperatura, a disponibilidade de água, o oxigênio, entre outros. Conhecer como se comporta a germinação de sementes com relação aos fatores de influência é de extrema importância, pois é por meio disso que o produtor pode buscar uma maior uniformidade da germinação, e conseqüentemente um aumento na produção com o menor gasto possível.

*Stevia rebaudiana* (Bert) Bertoni é o nome científico dado a planta herbácea paraguaia; a partir da qual os esteviol-glicosídeos foram extraídos por empresas industriais de adoçantes para a modificação de alimentos e bebidas (LEMUS-MONDACA et al., 2012; ARORA et al., 2010). Apesar das possibilidades de produção de plantas de colheita a partir do enraizamento das estacas, este método é frágil porque as plantas certificadas perdem a sua região meristemática onde os botões vegetativos se diferenciam em rebentos durante o estabelecimento da cultura e os transplantes atingidos morrem na primeira colheita. Assim, as sementes ainda são a melhor forma de se cultivar a cultura de *Stevia rebaudiana* sustentáveis em condições de campo. As sementes germinadas desenvolvem estruturas botânicas essenciais para a instalação de populações das plantas (NASSIF; VIEIRA; FERNADES, 1998; CARNEIRO, 2007). No entanto, altos níveis percentuais de mudas normais são afetadas pela luz, temperatura, água e oxigênio devido a fatores intrínsecos em embriões de sementes. Infelizmente, a heterogeneidade também rege dentro de ambientes controlados (POTVIN, 1993), bem como entre os germinadores de sementes, apesar de terem sido protegidos pelas condições da sala de germinação. Desta forma, o controle dos fatores que possam influenciar durante os testes de germinação de sementes e a produção de plantas tem sido a principal desvantagem

enfrentada pelos tecnólogos e plantadores de sementes. A germinação de sementes altas e uniformes da *Stevia rebaudiana* é importante para permitir projetos rentáveis de produção de colheitas, assim como eles formaram a referência para consultoria de agrônomos em condições de campo.

Por isso, a compreensão dos efeitos desses fatores de germinação é o primeiro passo antes do estabelecimento de áreas rentáveis. Pensando nisso, pesquisadores do Laboratório de Sementes do Departamento de Agronomia da Universidade Estadual de Maringá (UEM) realizaram um experimento sobre condições controladas para tentar entender as melhores condições que favorecem a sua germinação. O objetivo deles é entender se o tipo de luz utilizada e a posição que as sementes são alocadas dentro do germinadores podem influenciar na germinação das sementes. E com isso mapear sobre quais condições tem-se o maior número de plântulas normais, anormais e sementes não germinadas. Com o intuito de alcançar esse objetivo, realizou-se o experimento plantando 9.600 sementes de *Stevia rebaudiana* sobre condições controlada e ao fim registrou de acordo com cada um dos fatores de influência se as sementes plantadas geraram plântulas normais, anormais ou não germinaram, sendo neste caso, a variável resposta de interesse, uma variável aleatória qualitativa nominal.

Observa-se na literatura que quando se trata de análise referente a germinação de sementes de *Stevia rebaudiana* algumas das metodologias estatísticas empregadas são: análise descritiva juntamente com testes multivariados (ABDULLATEEF; OSMAN, 2011), análise de variância e/ou teste múltiplo de Duncan (SIMLAT et al., 2016; ÖZYİĞİT; ESRA; TURGUT, 2015; ABDULLATEEF; OSMAN; ZAINUDDIN, 2015; UÇAR; ÖZYİĞİT; TURGUT, 2016). Neste trabalho a metodologia empregada foi a análise de regressão para modelos lineares generalizados mistos. A justificativa de usar essa análise está relacionada com as características dos dados coletados, pois neste experimento observou-se variável resposta qualitativa e variáveis explicativas qualitativas com características fixa e aleatória. Sendo que uma variável explicativa será considerada fixa quando o pesquisador consegue controlar e reproduzir o experimento sobre as mesmas condições, como por exemplo a luz utilizada e a posição das sementes dentro do germinador. Já uma variável explicativa terá natureza aleatória quando o pesquisador não tem controle sobre ela, como por exemplo o lote que forneceu as sementes para o experimento, sendo este único e impossível de reprodução das mesmas sementes em um outro instante do tempo.

Os modelos lineares generalizados surgiram da necessidade em estudar variáveis respostas que apresentassem características diferentes da linear, e o mesmo foi proposto durante o século XX por Nelder e Wedderburn. A ideia dos autores era de ampliar o leque de opções em relação a distribuição da variável resposta, no qual é permitido modelar qualquer variável resposta que tenha como característica uma distribuição de probabilidade pertencente à família exponencial. Isso permitiu uma maior flexibilidade no estudo da relação funcional entre a média da variável resposta e o preditor linear. O avanço computacional impulsionado a partir da dé-



cada de 70 veio contribuir com a expansão e aplicação destes modelos, principalmente os que exigiam a utilização de processos iterativos na obtenção das estimativas de seus parâmetros (NELDER; BAKER, 1972; PAULA, 2004).

Porém, essa busca incessante por técnicas de modelagem estatísticas mais flexíveis e menos restritiva, fez com que Rigby e Stasinopoulos (2005) propusessem uma nova classe de modelos estatísticos de regressão que ficaram conhecidos por modelos aditivos generalizados para posição, escala e forma (GAMLSS), os quais possibilitam o ajuste de uma extensa família de distribuições para a variável resposta, saindo da restrição de pertencer a família exponencial. Além disso, esses modelos também permitem incorporar em sua estrutura de regressão variáveis explicativas que apresentam tanto efeitos fixos quanto efeitos aleatórios, dando assim maior flexibilidade e eficiência as análises estatística. Os ajustes desses modelos GAMLSS são obtidos através do ambiente estatístico R (R CORE TEAM, 2016), por meio do pacote “gamlss”.

A análise de regressão para modelos lineares generalizados mistos pode ser realizada tanto do ponto de vista frequentista quanto pelo ponto de vista bayesiana. E independentemente do enfoque dado, esse tipo de análise pode ser definida como uma metodologia estatística que permite explorar e inferir a relação existente entre a variável resposta/dependente com relação a uma ou mais variáveis explicativas/independentes.

A principal diferença entre o enfoque frequentista e o bayesiano é que na abordagem frequentista os dados são considerados aleatórios e os parâmetros quantidades fixa e desconhecida, já na abordagem bayesiana considera-se os dados e os parâmetros como quantidades aleatórias. O fato de considerar os parâmetros como quantidades aleatórias, do ponto de vista bayesiano, faz com que distribuições de probabilidade sejam atribuídas a eles possibilitando assim a inclusão de informações *a priori*, as quais podem ser informativas ou não informativas. Assim, a inferência bayesiana associa um modelo de probabilidade relacionado aos dados com a distribuição *a priori* dos parâmetros, que é aleatório, e por meio da distribuição condicional resume as informações dos dados observados com as dos parâmetros, dando origem a distribuição *a posteriori*. O avanço computacional impulsionou a utilização e desenvolvimento da inferência bayesiana, o que está relacionado a sua necessidade de programação computacional mas sofisticada na realização das simulações e conseqüentemente estimação dos parâmetros (BOLSTAD; CURRAN, 2016).

Na maioria dos casos a distribuição *a posteriori* não tem solução analítica, assim é necessário utilizar métodos iterativos. O método de Monte Carlo via Cadeias de Markov (MCMC) é um dos mais difundidos dentro da inferência bayesiana, sendo esse um método não determinístico. O elevado custo computacional exigido por esse método, fez com que pesquisadores buscassem alternativas para tornar a estimação da inferência bayesiana mais eficiente e rápida. Assim, uma alternativa que surgiu foi a Aproximação de Laplace Aninhada Integrada (INLA), do inglês, *Integrated Nested Laplace Approximation*, proposto por Rue, Martino e Chopin (2009), sendo esse um método determinístico, que combina aproximações de Laplace e inte-

gração numérica, tornando o mais rápido e eficiente. Os ajustes dos modelos com abordagem bayesiana, utilizando a metodologia INLA, podem ser obtidos através do ambiente estatístico R ([R CORE TEAM, 2016](#)), por meio do pacote “INLA”.

Assim, o objetivo deste trabalho foi apresentar uma modelagem estatística adequada referente aos dados hierárquicos sobre germinação de sementes de *Stevia rebaudiana* por meio das distribuições Binomial, Beta-binomial e Multinomial. E com isso identificar as melhores condições que possam vir a contribuir com a melhora da germinação destas sementes sob condições controladas.

---

## CAPÍTULO 2

---

# MATERIAIS E MÉTODOS

---

## 2.1 Materiais

Neste estudo foram utilizadas sementes de *Stevia rebaudiana* (Bert.) Bertoni cuja planta é importante do ponto de vista econômico, social, ambiental e político. Isso se deve à produção de glicosídeos tais como o esteviosídeo e o rebaudiosídeo extraídos das folhas. Além disso, estes compostos são não calóricos, anti-diabéticos, anti-glicêmicos, anti-cárie entre outras propriedades (RESENDE, 2000).



Figura 1 – Parcela de campo cultivada (a) e parcela em vaso cultivado (b) com *Stevia rebaudiana* (Bert.) Bertoni.

Com o objetivo de identificar as melhores condições para a germinação das sementes sob condições controladas, pesquisadores do Departamento de Agronomia da Universidade Estadual de Maringá realizaram em 2013 um experimento na Fazenda Experimental de Iguatemi na latitude de 23° 25' S; longitude de 51° 57' W e altitude de 506 m localizada no Distrito de Iguatemi, Município de Maringá, Paraná, Brasil, sendo as sementes de *Stevia rebaudiana* pertencente a 4 lotes (AGRICULTURA, 2009), obtidos durante o ano agrícola 2012/2013.

Neste experimento foram utilizados 4 germinadores, 2 tipos de iluminação (sob lâmpadas fluorescentes tipo luz do dia ou luz 1 e no escuro ou luz 2), sementes retiradas de 4 lotes distintos (AGRICULTURA, 2009), 96 bandejas de plástico branco com 100 células (CARNEIRO, 1996) de 4 cm<sup>3</sup> preenchidas com fibra de algodão farmacêutico umedecido com 2 mL de água destilada e deionizada (CARNEIRO, 2007) (CARNEIRO; GUEDES, 1992). Em cada um dos germinadores foram colocadas 24 bandejas distribuídas aleatoriamente da seguinte forma: 8 na parte superior, 8 na parte média e 8 na parte inferior de cada um dos germinadores; destas 8 bandejas 4 eram replicações de um lote de sementes e as outras 4 compunham o outro lote; assim cada germinador só tinha sementes pertencentes a dois lotes distintos, conforme pode-se observar na Figura 2, que é o esquema geral referente ao delineamento do experimento.

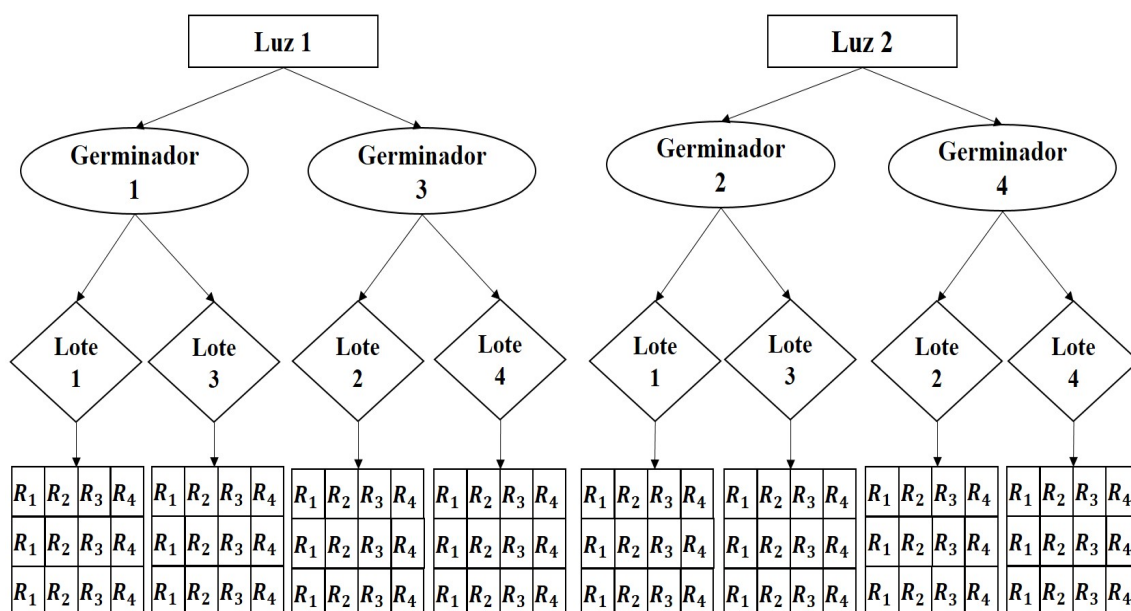


Figura 2 – Delineamento do experimento realizado com as sementes de *Stevia rebaudiana*.

Conforme a descrição do experimento na Figura 2, tem-se na Tabela 1 um resumo das variáveis explicativas utilizadas nesse estudo.

Tabela 1 – Descrição das variáveis explicativas utilizadas no experimento.

Variável	Descrição	Níveis
Lote	O lote da qual as sementes foram retiradas para o experimento	4
Luz	O tipo de luz que foi utilizada na germinação das sementes	2
Posição	A posição que as sementes foram alocadas dentro do germinador	3
Germinador	Os germinadores utilizados para a realização do experimento	4

O experimento durou 10 dias, momento no qual foi feito o registro do que ocorreu com cada semente, ou seja, se gerou plântulas normais, plântulas anormais ou se não germinou (ISTA, 2013; AGRICULTURA, 2009) (Figura 3), sendo essa a variável resposta.

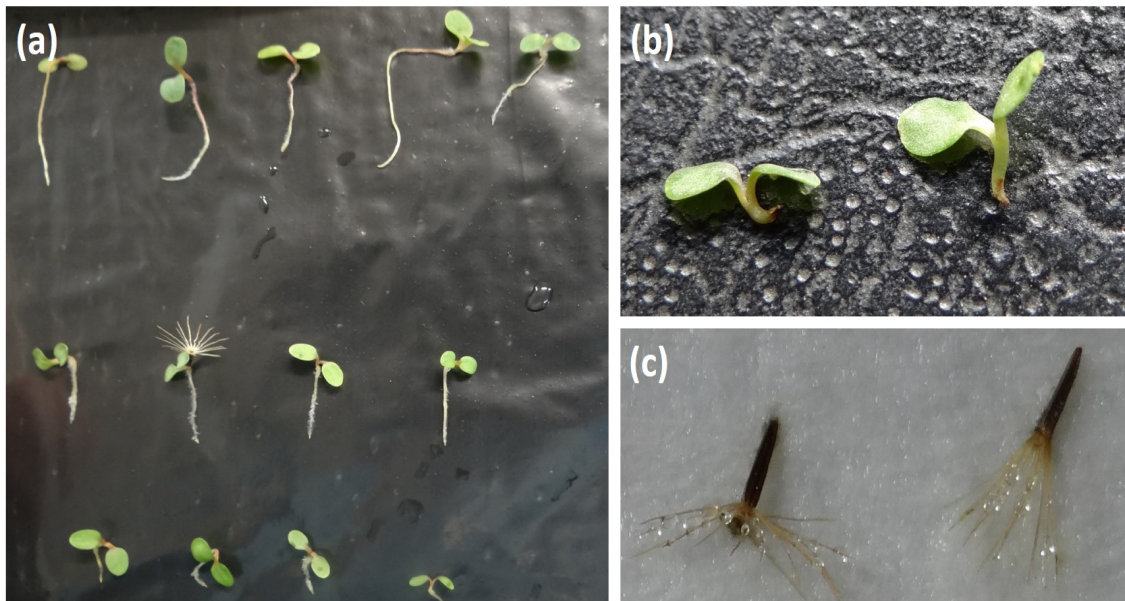


Figura 3 – Características morfo-anatômicas das plântulas normais (a), plântulas anormais (b) e sementes não germinadas (c).

## 2.2 Métodos

### 2.2.1 Modelos de regressão linear generalizado misto

Para modelar o conjunto de dados obtidos no experimento apresentado na Figura 2 utilizou-se modelos lineares generalizados mistos, sendo estes modelos estatísticos que descrevem o comportamento da variável resposta com relação a uma ou mais variáveis explicativas, as quais podem envolver tanto efeitos fixos como aleatórios. Associada a essa variável resposta, há uma distribuição de probabilidade que busca relacionar os valores observados com as suas respectivas probabilidades de ocorrência. Nesse sentido, esta seção tem o objetivo de descrever as distribuições Binomial, Beta-binomial e Multinomial com as suas respectivas estruturas de regressão.

#### 2.2.1.1 Modelo de regressão Binomial misto

A distribuição de Bernoulli, caso especial da Binomial assumi apenas duas respostas possíveis, como sucesso ou fracasso, em que atribui-se o valor 1 ao sucesso com probabilidade  $\pi$  de ocorrer e o valor 0 ao fracasso com probabilidade de ocorrência  $1 - \pi$ . A sua função de probabilidade é dada por:

$$P(Y = y) = \pi^y (1 - \pi)^{(1-y)}, y = 0 \text{ ou } 1, \quad (2.1)$$

na qual, a variável aleatória  $Y_i \sim \text{Ber}(\pi)$  tem média e variância dada por  $E(Y) = \pi$  e  $\text{Var}(Y) = \pi(1 - \pi)$ .

Após a repetição de  $n$  ensaios independentes de Bernoulli, em que todas as ocorrências têm a mesma probabilidade de sucesso  $\pi$ , tem-se a distribuição Binomial. A variável aleatória agora passa a ser o número total de sucessos observados, representado aqui por  $y$ , em  $n$  ensaios de Bernoulli (MAGALHÃES, 2006). Com notação  $Y \sim \text{Bin}(n, \pi)$ , a sua função de probabilidade tem dois parâmetros  $n$  conhecido e  $\pi$  desconhecido, e é dada por:

$$P(Y = y) = \binom{n}{y} \pi^y (1 - \pi)^{(n-y)}, \quad y = 0, 1, 2, \dots, n. \quad (2.2)$$

Na Figura 4 é apresentado o comportamento da função de probabilidade da distribuição Binomial para  $n = 100$  e três valores distintos de  $\pi$ , sendo eles 0,2; 0,5 e 0,8.

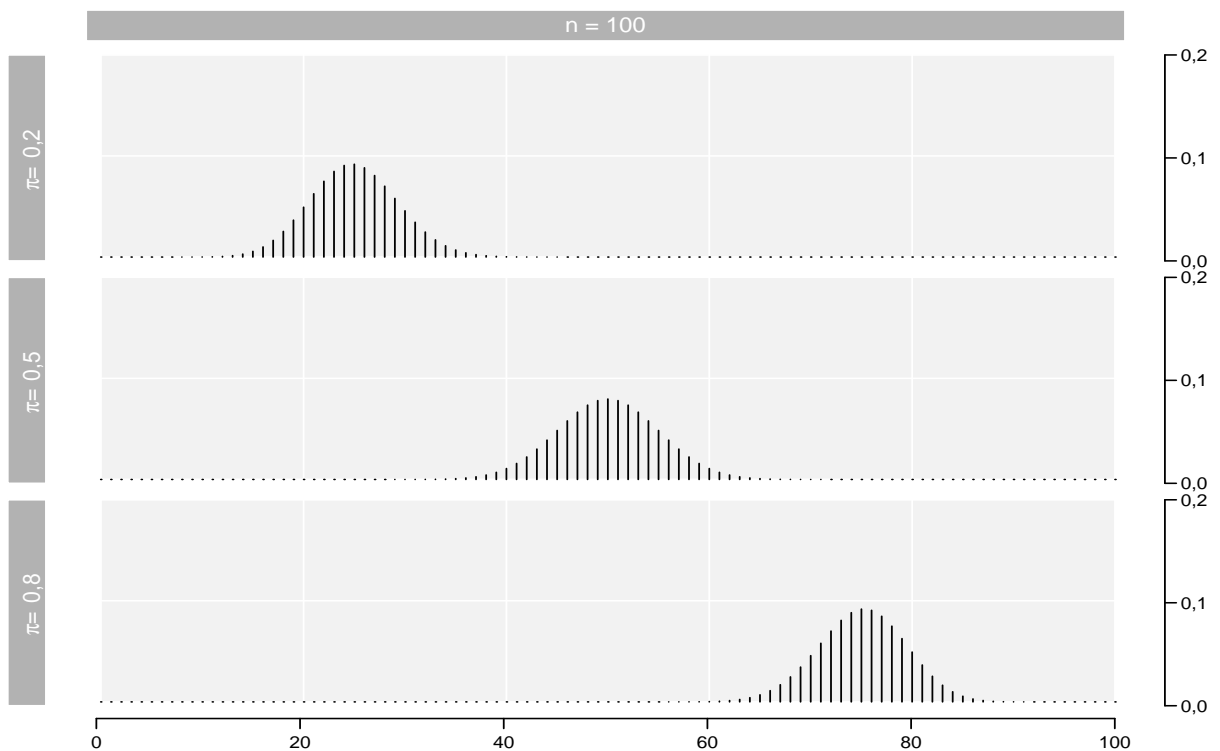


Figura 4 – Função distribuição de probabilidade da Binomial para valores de  $\pi$  iguais a 0,2; 0,5; 0,8 e  $n = 100$ .

A função de distribuição acumulada da distribuição Binomial é calculada por:

$$F_Y(Y = y) = \begin{cases} 0 & , \text{ se } y < 0; \\ \sum_{k=0}^y \binom{n}{k} \pi^k (1 - \pi)^{(n-k)} & , \text{ se } 0 \leq y \leq n; \\ 1 & , \text{ se } y > n. \end{cases}$$

A média e a variância de uma variável aleatória  $Y$  com distribuição Binomial é dada, respectivamente, por:

$$E(Y) = n\pi$$



e

$$\text{Var}(Y) = n\pi(1 - \pi).$$

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 com efeitos mistos na forma matricial é dada por:

$$g(\boldsymbol{\pi}) = \boldsymbol{\eta} = \underbrace{\mathbf{X}\boldsymbol{\beta}}_{\text{Fixo}} + \underbrace{\mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}}_{\text{Aleatório}}, \quad (2.3)$$

em que;

$g(\boldsymbol{\pi})$  é a função de ligação que relaciona o vetor de parâmetros  $\boldsymbol{\pi}$  da distribuição  $\text{Bin}(n, \boldsymbol{\pi})$  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 referente aos efeitos fixos;

$\mathbf{b}$  é o vetor  $q \times 1$  de parâmetros referente aos efeitos aleatórios, em que assume-se com distribuição Normal multivariada  $N(\mathbf{0}, \mathbf{D})$ , com vetor de médias igual a zero e matriz  $\mathbf{D}$  de covariância dos efeitos aleatórios com dimensão  $q \times q$ ;

$\boldsymbol{\epsilon}$  é um vetor  $n \times 1$  de erros aleatórios associado ao modelo com distribuição multivariada  $N(\mathbf{0}, \mathbf{R})$ , sendo  $\mathbf{R}$  a matriz de covariância dos erros com dimensão  $n \times n$ .

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

$$E(Y_i) = ng^{-1}(\eta_i)$$

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, descrito na seção 2.2.1.4.

### 2.2.1.2 Modelo de regressão Beta-binomial misto

A distribuição discreta Beta-binomial surgiu da distribuição Binomial com parâmetros  $n$  e  $\pi$ , em que  $\pi$  tem distribuição Beta. Essa distribuição leva em consideração a possibilidade das unidades amostrais não terem as mesmas probabilidades de sucesso, isso possibilita modelar de forma adequada dados que apresentam subdispersão ou sobredispersão, quando os dados não apresentam nenhuma destas duas características os resultados com a distribuição Beta-binomial se aproximam dos resultados obtidos com a distribuição Binomial. A distribuição Binomial considera que a probabilidade de sucesso em todas as unidades amostrais seja constante.

A função de probabilidade da variável aleatória  $Y$  é determinada pela densidade conjunta, entre a distribuição Binomial com a Beta, em todo o espaço paramétrico de  $\pi$ , ou seja,  $\pi$  pertencendo ao intervalo  $(0, 1)$ . A função de probabilidade da distribuição Beta-binomial, denotada aqui por  $BB(n, \pi, \sigma)$  é dada por (RIGBY; STASINOPOULOS, 2012):

$$P(y|n, \pi, \sigma) = \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma\left(\frac{1}{\sigma}\right) \Gamma\left(y + \frac{\pi}{\sigma}\right) \Gamma\left(n + \frac{(1-\pi)}{\sigma} - y\right)}{\Gamma\left(n + \frac{1}{\sigma}\right) \Gamma\left(\frac{\pi}{\sigma}\right) \Gamma\left(\frac{1-\pi}{\sigma}\right)} \quad (2.4)$$

para  $y = 0, 1, \dots, n$  em que  $0 < \pi < 1$  e  $\sigma > 0$ , sendo  $n$  um número inteiro positivo conhecido.

Na Figura 5, é apresentado o comportamento da função de probabilidade da distribuição Beta-binomial para valores de  $\pi$  iguais a 0,1; 0,3; 0,5; 0,7 e 0,9 e valores de  $\sigma$  iguais a 0,01; 0,10; 0,50 e 1,00.

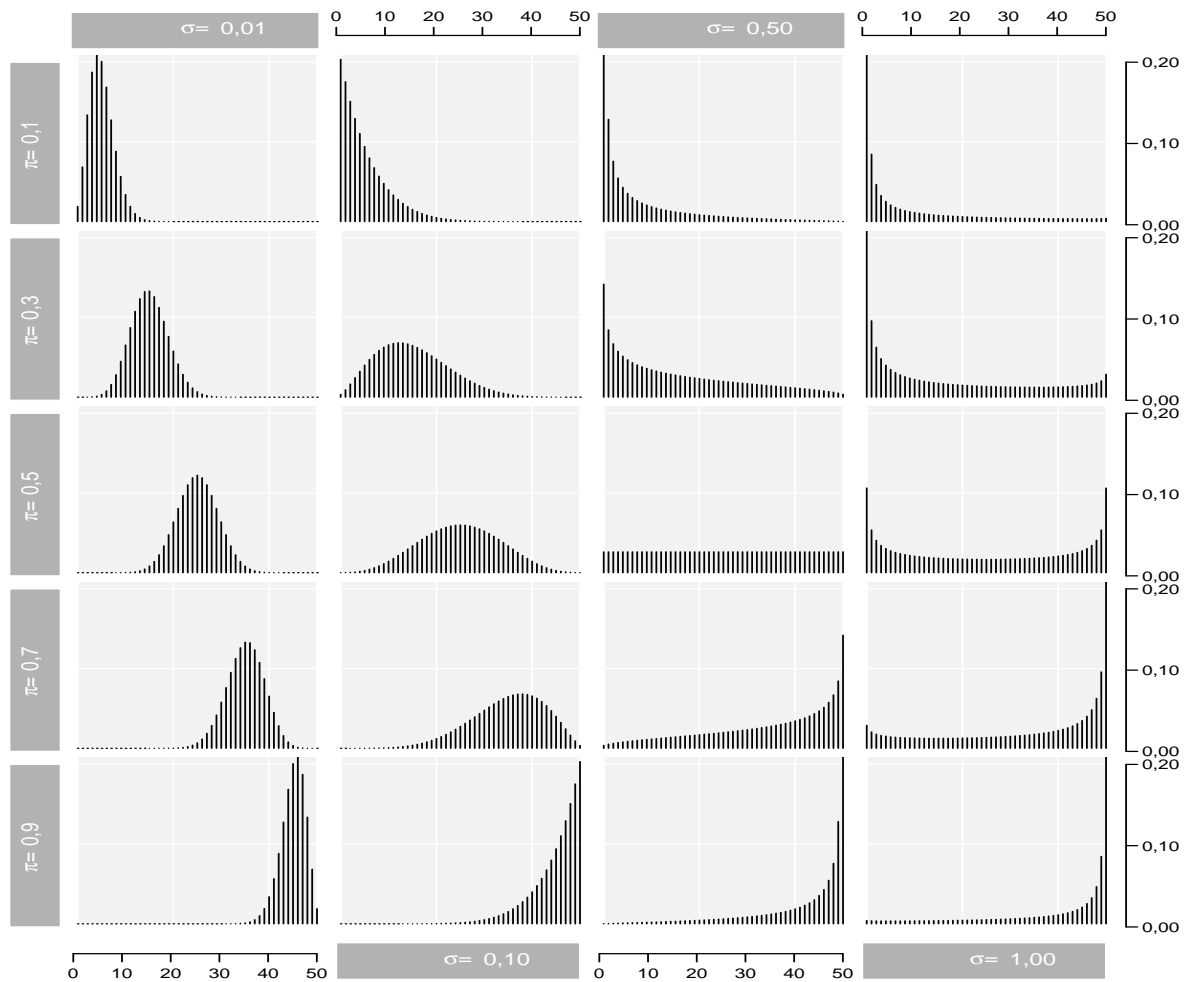


Figura 5 – Função distribuição de probabilidade da Beta-binomial com valores de  $\pi$  iguais a 0,1; 0,3; 0,5; 0,7 e 0,9 e valores de  $\sigma$  iguais a 0,01; 0,10; 0,50 e 1,00.

A média e a variância de uma variável aleatória  $Y$  com distribuição Beta-binomial é determinada, respectivamente, por (RIGBY; STASINOPOULOS, 2012):

$$E(Y|n, \pi, \sigma) = n\pi$$



e

$$\text{Var}(Y|n, \pi, \sigma) = n\pi(1 - \pi) \left[ 1 + \frac{\sigma}{1 + \sigma}(n - 1) \right].$$

Sendo  $Y_1, Y_2, \dots, Y_n$  uma amostra aleatória independente da variável aleatória  $Y$ , em que  $Y_i \sim \text{BB}(n_i, \pi_i, \sigma_i)$ , o modelo de regressão linear generalizado com efeitos mistos na forma matricial é dada por:

$$g_1(\boldsymbol{\pi}) = \boldsymbol{\eta}_1 = \underbrace{\mathbf{X}\boldsymbol{\beta}_1}_{\text{Fixo}} + \underbrace{\mathbf{Z}\mathbf{b}_1}_{\text{Aleatório}} + \boldsymbol{\epsilon}, \quad (2.5)$$

e

$$g_2(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = \underbrace{\mathbf{X}\boldsymbol{\beta}_2}_{\text{Fixo}} + \underbrace{\mathbf{Z}\mathbf{b}_2}_{\text{Aleatório}} + \boldsymbol{\epsilon}, \quad (2.6)$$

em que;

$g_1(\boldsymbol{\pi})$  e  $g_2(\boldsymbol{\sigma})$  são as funções de ligação que relacionam, respectivamente, os vetores de parâmetros  $\boldsymbol{\pi}$  e  $\boldsymbol{\sigma}$  da distribuição  $\text{BB}(\mathbf{n}, \boldsymbol{\pi}, \boldsymbol{\sigma})$  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 referente aos efeitos fixos;

$\mathbf{b}$  é o vetor  $q \times 1$  de parâmetros referente aos efeitos aleatórios, em que assume-se com distribuição multivariada  $N(\mathbf{0}, \mathbf{D})$ , com vetor de médias igual a zero e matriz  $\mathbf{D}$  de covariância dos efeitos aleatórios com dimensão  $q \times q$ ;

$\boldsymbol{\epsilon}$  é um vetor  $n \times 1$  de erros aleatórios associado ao modelo com distribuição multivariada  $N(\mathbf{0}, \mathbf{R})$ , em que  $\mathbf{R}$  é a matriz de covariância associada aos erros com dimensão  $n \times n$ .

Sabendo que,  $g_1(\pi_i)$  e  $g_2(\sigma_i)$  representam a função de ligação adequada, utilizada para relacionar os parâmetros  $\pi_i$  e  $\sigma_i$  ao seu, respectivo preditor linear  $\eta_{1i}$  e  $\eta_{2i}$ , tem-se que a média e variância da variável aleatória  $Y_i$  após a regressão são dadas, respectivamente, por (RIGBY; STASINOPOULOS, 2012):

$$E(Y_i|n, \pi, \sigma) = ng_1^{-1}(\eta_{1i})$$

e

$$\text{Var}(Y_i|n, \pi, \sigma) = ng_1^{-1}(\eta_{1i})(1 - g_1^{-1}(\eta_{1i})) \left[ 1 + \frac{g_2^{-1}(\eta_{2i})}{1 + g_2^{-1}(\eta_{2i})}(n - 1) \right].$$

Desta forma, pode-se utilizar o Método Delta, descrito na seção 2.2.1.4, para obter  $E(Y_i)$  e  $\text{Var}(Y_i)$  e assim encontrar as estimativas intervalares para um valor predito de  $Y_i$ .

### 2.2.1.3 Modelo de regressão Multinomial misto

A distribuição Multinomial é uma generalização da distribuição Binomial, nos casos em que a variável aleatória pode assumir mais do que dois resultados possíveis. Imagine um experimento aleatório com  $k$  eventos mutuamente exclusivos,  $A_1, A_2, A_3, \dots, A_k$ , ou seja, esses eventos formam um espaço amostral do experimento, no qual, as possibilidades de

sucesso é dada por  $P(A_i) = \pi_i, i = 1, 2, \dots, k$ . Considere agora o mesmo experimento com  $n$  tentativas independentes, sendo que os  $\pi_i, i = 1, 2, \dots, k$  permanecem constantes durante as repetições, com  $\sum_{i=1}^k \pi_i = 1$ . Em que a variável aleatória  $\mathbf{Y}$  formada pelo vetor  $(Y_1, Y_2, \dots, Y_k)$ , contém os números de ocorrências dos eventos  $A_1, A_2, A_3, \dots, A_k$ , respectivamente, com  $\sum_{i=1}^k Y_i = n$  (MORETTIN, 2010). Nestas condições, tem-se a função de probabilidade conjunta da distribuição Multinomial, com notação  $\text{Multi} \sim (n, \pi_1, \dots, \pi_k)$ , é dada por:

$$P(Y_1 = n_1, Y_2 = n_2, \dots, Y_k = n_k) = \frac{n!}{n_1! n_2! \dots n_k!} \pi_1^{n_1} \pi_2^{n_2} \dots \pi_k^{n_k},$$

Na Figura 6 é apresentado o comportamento da função de probabilidade conjunta da distribuição Multinomial para os casos em que  $k = 3$ . Como na distribuição Multinomial  $\sum_{i=1}^k \pi_i = 1$ , fixando  $\pi_1$  e  $\pi_2$  o valor de  $\pi_3$  é conhecido, ou seja,  $\pi_3 = 1 - (\pi_1 + \pi_2)$ .

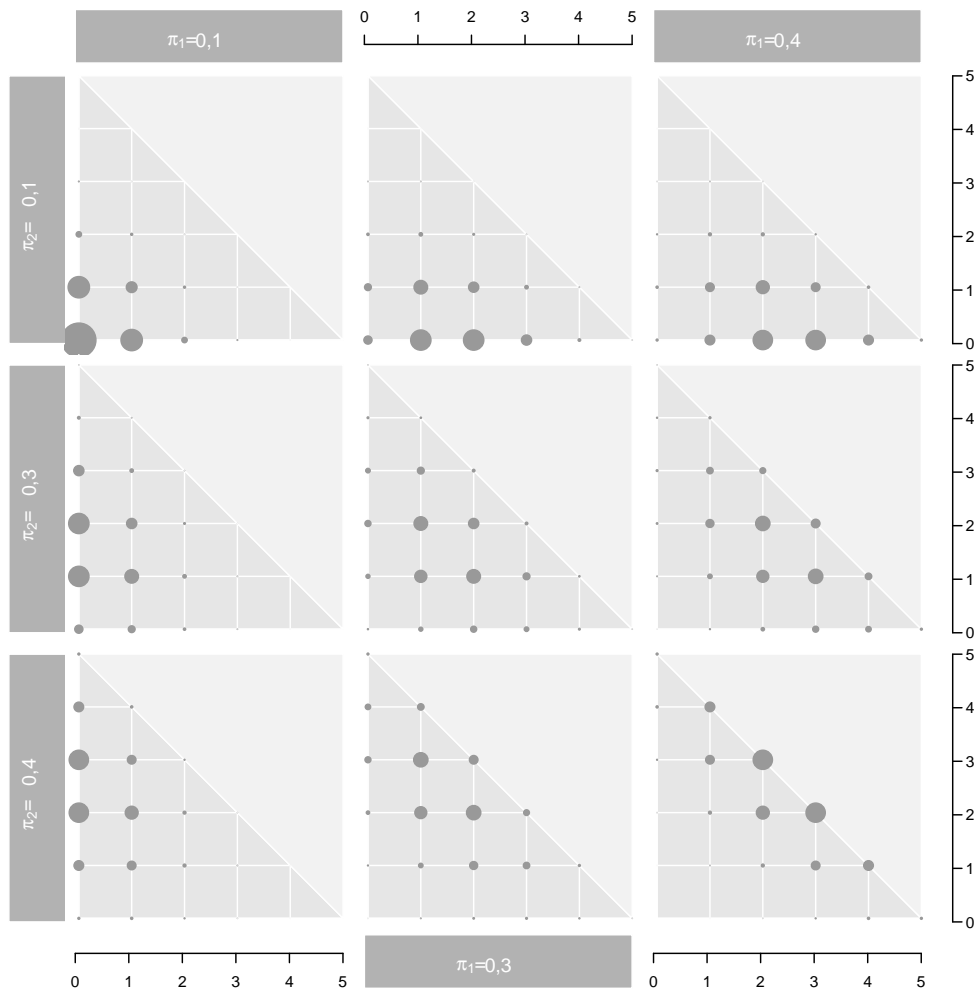


Figura 6 – Função distribuição de probabilidade da Multinomial para  $\pi_1$  e  $\pi_2$  iguais a 0, 1; 0, 3 e 0, 4, sendo que a dimensão da bola representa a função de densidade nos pontos, para  $n = 5$ .

A variável aleatória  $Y$  com distribuição Multinomial tem média dada por:

$$E(Y) = n\boldsymbol{\pi},$$

em que  $\boldsymbol{\pi}$  é um vetor  $k \times 1$  definido por  $\boldsymbol{\pi} = [\pi_1, \pi_2, \dots, \pi_k]^T$ , e variância

$$\text{Var}(Y) = n \boldsymbol{\Sigma},$$

sendo  $\boldsymbol{\Sigma}$  uma matriz de covariância  $k \times k$ , cujos elementos são

$$\boldsymbol{\Sigma}_{ij} = \begin{cases} \pi_i(1 - \pi_i) & , \text{ se } j = i \\ -\pi_i\pi_j & , \text{ se } j \neq i. \end{cases}$$

Se  $Y_1, Y_2, \dots, Y_k$  segue distribuição Multinomial com parâmetros  $n$  e  $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_k)$ , então cada um dos  $Y_i \sim \text{Bin}(n_i, \pi_i)$ , ou seja, a distribuição marginal da Multinomial tem distribuição Binomial.

Sendo  $Y_1, Y_2, \dots, Y_n$  uma amostra aleatória independente de uma variável aleatória  $Y$ , tal que  $Y_i \sim \text{Multi}(n_i, \pi_1, \dots, \pi_k)$ , o modelo de regressão linear generalizado com efeitos mistos na forma matricial é dada por:

$$g_1(\pi_1) = \boldsymbol{\eta}_1 = \underbrace{\mathbf{X}\boldsymbol{\beta}_1}_{\text{Fixo}} + \underbrace{\mathbf{Z}\mathbf{b}_1}_{\text{Aleatório}} + \boldsymbol{\epsilon} \quad (2.7)$$

e

$$g_2(\pi_2) = \boldsymbol{\eta}_2 = \underbrace{\mathbf{X}\boldsymbol{\beta}_2}_{\text{Fixo}} + \underbrace{\mathbf{Z}\mathbf{b}_2}_{\text{Aleatório}} + \boldsymbol{\epsilon}, \quad (2.8)$$

em que;

$g_1(\pi_1)$  e  $g_2(\pi_2)$  são as funções de ligação que relacionam, respectivamente, os vetores de parâmetros  $\boldsymbol{\pi}_1$  e  $\boldsymbol{\pi}_2$  da distribuição Multinomial, com os seus respectivos preditores lineares  $\boldsymbol{\eta}_1$  e  $\boldsymbol{\eta}_2$ ;

$\mathbf{X}$  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 referente aos efeitos fixos;

$\mathbf{b}$  é o vetor  $q \times 1$  de parâmetros referente aos efeitos aleatórios, assume-se com distribuição multivariada  $N(\mathbf{0}, \mathbf{D})$ , com vetor de médias igual a zero e matriz  $\mathbf{D}$  de covariância dos efeitos aleatórios com dimensão  $q \times q$ ;

$\boldsymbol{\epsilon}$  é um vetor  $n \times 1$  de erros aleatórios associado ao modelo com distribuição multivariada  $N(\mathbf{0}, \mathbf{R})$ , em que  $\mathbf{R}$  representa a matriz de covariância dos erros.

Sabendo que,  $g_1(\pi_{1i})$  e  $g_2(\pi_{2i})$  representam as funções de ligações adequadas, utilizadas para relacionarem os parâmetros  $\pi_{1i}$  e  $\pi_{2i}$  ao seu respectivo preditor linear  $\eta_{1i}$  e  $\eta_{2i}$ , tem-se

que a média e variância de um dado valor de  $Y_i$  após a regressão são dadas, respectivamente, por (MORETTIN, 2010):

$$E(Y_i) = ng_{1i}^{-1}(\eta_{1i})g_{2i}^{-1}(\eta_{2i})\left(1 - g_1^{-1}(\eta_{1i}) - g_{2i}^{-1}(\eta_{2i})\right)$$

e

$$Var(Y_i) = n \Sigma,$$

e que  $\Sigma$  é uma matriz de covariância  $k \times k$ , cuja as entradas são

$$\Sigma_{ij} = \begin{cases} g_i^{-1}(\eta_i)(1 - g_i^{-1}(\eta_i)) & , \text{ se } j = i \\ -g_i^{-1}(\eta_i)g_j^{-1}(\eta_j) & , \text{ se } j \neq i. \end{cases}$$

Assim, por meio do Método Delta, descrito na seção 2.2.1.4, pode se obter a  $E(Y_i)$  e  $Var(Y_i)$ , e assim construir os intervalos de confiança para valores preditos de  $Y_i$ .

#### 2.2.1.4 Método Delta

O Método Delta é utilizado para aproximar um vetor aleatório por meio de uma expansão em série de Taylor, sendo um método simples porém útil, para deduzir a distribuição assintótica de variáveis e assim estimar a média e variância de uma dada função que possuem parâmetros estimados por máxima verossimilhança. Os resultados obtidos pela aproximação são garantidos pela propriedade de invariância dos estimadores de máxima verossimilhança.

**Teorema:** Seja  $Y_n$  uma sequência de variáveis aleatórias que satisfaz

$$\sqrt{n}(Y_n - \theta) \rightarrow N(0, \sigma^2).$$

Para uma dada função  $g$  e um específico valor de  $\theta$ , suponha que  $g'(\theta)$  existe e é diferente de zero. Então

$$\sqrt{n}(g(Y_n) - g(\theta)) \rightarrow N(0, \sigma^2[g'(\theta)]^2).$$

*Demonstração:* A expansão de  $g(Y_n)$  em série de Taylor de ordem 1 em torno de  $\theta$  é

$$g(Y_n) = g(\theta) + g'(\theta)(Y_n - \theta) + \epsilon$$

onde o  $\epsilon \rightarrow 0$  quando  $Y_n \rightarrow \theta$ . Aplicando o Teorema de Slutsky, que está descrito em Casella e Berger (2002), é possível afirmar que

$$\sqrt{n}(g(Y_n) - g(\theta)) = g'(\theta)\sqrt{n}(Y_n - \theta).$$

Desta forma, tem-se que para o caso univariado

$$g(\hat{\theta}) \rightarrow N\left(g(\theta), [g'(\theta)]^2 Var(\hat{\theta})\right).$$

Informações e demonstrações sobre o método delta aplicado a funções multivariadas estão disponíveis em Casella e Berger (2002) e Lehmann e Casella (2006).

## 2.2.2 Abordagem Frequentista

Para ajustar os modelos propostos, do ponto de vista frequentista, utilizou-se o pacote “gamlss” do ambiente estatístico R (R CORE TEAM, 2016). Desta forma, essa seção apresentará as ferramentas utilizadas pelo pacote para estimar, selecionar e validar o modelo de regressão em estudo.

### 2.2.2.1 Modelo aditivo generalizado para locação, escala e forma

O modelo aditivo generalizado para locação, escala e forma (GAMLSS) foi proposto por Rigby e Stasinopoulos (2005), como uma forma alternativa para driblar algumas restrições impostas pelos modelos lineares e lineares generalizados, que passaram a ser agora um caso particular do modelo aditivo generalizado para locação, escala e forma, conforme representado na Figura 7.

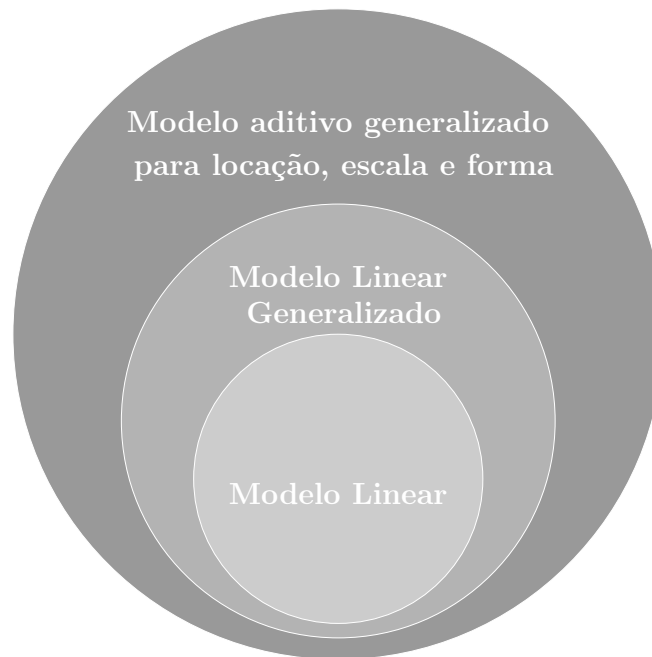


Figura 7 – Representação de como estão alocados os modelos de regressão.

O modelo GAMLSS substitui a restrição de que a variável resposta tenha que pertencer a família exponencial, por uma classe de família com distribuição mais geral, designada aqui por  $D$ . Essa distribuição  $D(y|\theta)$ , em que  $\theta$  representa um vetor de parâmetros, e  $y$  a variável resposta que pode ter associada qualquer distribuição contínua ou discreta. O pacote “gamlss” permite o ajuste de funções de distribuições com até quatro parâmetros. Nos casos em que a distribuição tenha mais de quatro parâmetros a metodologia continua sendo válida, porém a programação para a análise ainda não está implementada. O vetor de parâmetros  $\theta = (\mu, \sigma, \nu, \tau)$  é usualmente caracterizados em posição ( $\mu$ ), escala ( $\sigma$ ), assimetria ( $\nu$ ) e curtose ( $\tau$ ). (RIGBY; STASINOPOULOS, 2012).

Assim, o modelo GAMLSS assume que dado um vetor  $\mathbf{Y} = (y_1, y_2, \dots, y_n)^t$  de  $n$  observações independentes e identicamente distribuídas, tem-se:

$$g_k(\theta_k) = \boldsymbol{\eta}_k = \underbrace{\mathbf{X}\boldsymbol{\beta}_k}_{\text{Fixo}} + \underbrace{\mathbf{Z}\mathbf{b}_k}_{\text{Aleatório}} + \boldsymbol{\epsilon} \quad (2.9)$$

em que;

$g_k(\theta_k)$  é a função de ligação adequada para os parâmetros de interesse, que têm a função de relacionar esse parâmetro ao seu respectivo preditor linear  $\eta_k$ , sendo  $k = 1, 2, 3$  ou 4.

$\mathbf{X}$  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  $p \times 1$  de parâmetros referente aos efeitos fixos;

$\mathbf{b}$  é o vetor  $q \times 1$  de parâmetros referente aos efeitos aleatórios, em que assume-se com distribuição multivariada  $N(\mathbf{0}, \mathbf{D})$ , com vetor de médias igual a zero e matriz  $\mathbf{D}$  de covariância dos efeitos aleatórios com dimensão  $q \times q$ ;

$\boldsymbol{\epsilon}$  é um vetor  $n \times 1$  de erros aleatórios associado ao modelo com distribuição multivariada  $N(\mathbf{0}, \mathbf{R})$ , em que  $\mathbf{R}$  representa a matriz de covariância dos erros com dimensão  $n \times n$ .

A função de ligação, representada aqui por  $g_k(\theta_k)$ , é uma função monótona e diferenciável que tem o objetivo de relacionar a média ao preditor linear,  $\eta$ . Na Tabela 2, são apresentadas algumas funções de ligação usuais utilizadas de acordo com o suporte do parâmetro. O pacote “gamlss” disponibiliza na função *gamlss* os argumentos *mu.link*, *sigma.link*, *nu.link* e *tau.link* para que o usuário possa definir a função de ligação desejada referente a cada um dos parâmetros de interesse. Disponibiliza também a função *make.link.gamlss()* a qual possibilita a criação de uma nova função de ligação, ainda não implementada e a função *show.link()* que tem o papel de mostrar a função de ligação utilizada.

Tabela 2 – Funções de ligação.

Função de ligação	$g(\theta_k) = \eta$	$\theta_k = g^{-1}(\eta)$	Suporte
Identidade	$\theta_k$	$\theta_k$	$(-\infty, +\infty)$
Logística	$\log\left(\frac{\theta_k}{1-\theta_k}\right)$	$\frac{\exp(\eta)}{1+\exp(\eta)}$	$(0, 1)$
Logarítmica	$\log(\theta_k)$	$\exp(\eta)$	$(0, +\infty)$
Probit	$\Phi^{-1}(\theta_k)$	$\Phi(\eta)$	$(0, 1)$
Log-log	$-\log(-\log(\theta_k))$	$\exp(-\exp(-\eta))$	$(0, 1)$

\*  $\Phi$  é a função da distribuição acumulada da distribuição Normal padrão.

#### 2.2.2.1.1 Método de estimação

Os modelos estatísticos possuem parâmetros cujos verdadeiros valores são desconhecidos. Assim é necessário utilizar algum método de estimação, que baseado nas informações

contida na amostra forneça uma estimativa que seja o mais próximo possível do verdadeiro valor dos parâmetros populacionais. A estimação dos parâmetros considerando o modelo (2.9), será obtida aqui pela maximização da função da verossimilhança penalizada. Sobre a suposição de que todas as observações da variável resposta são independentes e identicamente distribuídas tem-se que a função log-verossimilhança penalizada do modelo GAMLSS (2.9) é obtida por:

$$l_p = l - \frac{1}{2} \sum_{k=1}^4 \sum_{j=1}^{j_k} \lambda_{kj} b_{kj}^T G_{kj} b_{kj}, \quad (2.10)$$

em que  $l$  é a função log-verossimilhança definida como  $l = \sum_{i=1}^n \log f(y_i | \mu_i, \sigma_i, \nu_i, \tau_i)$ , sendo  $f(\cdot)$  a função de densidade de probabilidade da variável resposta e  $\boldsymbol{\lambda}$  é o vetor de hiperparâmetros fixos de suavização,  $\mathbf{b}$  são as medidas que representam os parâmetros dos efeitos aleatórios a serem estimados e  $\mathbf{G}$  é a matriz simétrica de dimensão  $q_k \times q_k$  que pode depender dos vetores de hiperparâmetros, no qual a  $\mathbf{G}^{-1}$  é a matriz de covariância.

Devido à dificuldade ou não existência de solução analítica, para encontrar o estimador desses parâmetros foi implementado no ambiente R via pacote “gamlss” os dois algoritmos Cole e Green (CG) e Rigby e Stasinopoulos (RS), os quais serão utilizados na estimação da função (2.10):

- Algoritmo CG: É uma generalização do algoritmo de Cole e Green (1992), no qual utiliza informações esperada ou aproximada da primeira ou segunda derivada transversal da função de log-verossimilhança com relação aos parâmetros  $\boldsymbol{\theta}$  da função de distribuição, caso essa função de distribuição tenha quatro parâmetros por exemplo  $\boldsymbol{\theta} = (\mu, \sigma, \nu, \tau)$ . No entanto, quando os parâmetros da funções de densidade de probabilidade,  $f(y|\boldsymbol{\theta})$  são ortogonais, isto é, os valores esperados da matriz de derivadas de segunda ordem de  $l$  em (2.10) fora da diagonal principal são zeros, um algoritmo de estimação mais simples com o algoritmo RS é utilizado. É valido destacar também, que no caso totalmente paramétrico, o algoritmo CG corresponde ao Método Escore de Fisher;
- Algoritmo RS: É uma generalização do algoritmo usado por Rigby e Stasinopoulos (1996) com o objetivo de ajustar a média e a dispersão dos modelos aditivos. E nesse caso, as derivadas da diagonal principal da função log-verossimilhança não são requeridas.

Mais detalhes sobre a estimação dos parâmetros e sobre os algoritmos utilizados são encontrados em Rigby e Stasinopoulos (2005), Rigby e Stasinopoulos (2012) e no Pawitan (2001).

### 2.2.2.1.2 Seleção do modelo

A seleção de um modelo final adequado, que será adotado, constitui uma parte muito importante na modelagem estatística, e é por meio dessa seleção que se busca um modelo capaz de explicar de forma satisfatória a variável resposta com relação às variáveis explicativas. Além disso espera-se que o modelo escolhido apresente o menor erro possível. Para auxiliar no processo de construção de um modelo utilizou-se o desvio global ajustado (Deviance Global) e o Critério de Informação de Akaike Generalizado (GAIC), sendo que o menor valor destas medidas evidencia um modelo com menor erro.

- Deviance Global: É definida por:

$$DG = -2l(\hat{\theta})$$

sendo  $l(\hat{\theta}) = \sum_{i=1}^n l(\hat{\theta}^i)$  a função log-verossimilhança.

- GAIC: É obtido por meio da Deviance Global somada a uma penalidade fixa  $\omega$  para cada grau de liberdade efetivo que é usado no modelo, ou seja:

$$GAIC(\omega) = DG + \omega df,$$

em que,  $\omega df$  representa o total de graus de liberdade efetivos utilizados no modelo e a DG. Quando  $\omega = 2$  tem-se o Critério de Informação de Akaike (AIC) original e quando  $\omega = \log(n)$  o Critério bayesiano de Schwarz (BIC) original. Tanto o critério de AIC como o BIC, permitem comparar modelos não-encaixados, no qual eles penalizam os modelos com maiores números de parâmetros.

O modelo selecionado será o que apresentar a menor DG e o menor  $GAIC(\omega)$ . Para um estudo mais aprofundado sobre essas medidas de seleção de modelos recomenda-se a consulta dos guias do pacote “gamlss” (RIGBY; STASINOPOULOS, 2012).

### 2.2.2.1.3 Diagnóstico

Para o diagnóstico dos resíduos o pacote “gamlss” utiliza os resíduos quantílicos aleatorizados, proposto por Dunn e Smyth (1996). A vantagem destes resíduos é que para ajustes adequados, esses resíduos tem distribuição Normal padrão. E isso é válido para variável resposta com qualquer distribuição de probabilidade.

Dado que uma distribuição  $f(y|\theta)$  é ajustada para valores observados de  $y_i$ , com  $i = 1, 2, \dots, n$ , tem-se que os resíduos quantílicos aleatorizados são dados por:

$$\hat{r}_i = \Phi^{-1}(\hat{u}_i), \quad (2.11)$$



em que  $\Phi^{-1}$  é a função inversa da distribuição acumulada de uma variável com distribuição Normal padrão. E os  $\hat{u}_i$ 's são os resíduos quantílicos, os quais são definidos de forma diferente para as variáveis respostas contínuas e discretas.

Quando a variável resposta  $Y_i$  tem distribuição contínua,  $\hat{u}_i = F(y_i|\hat{\theta}_i)$ , em que  $F(y_i|\hat{\theta}_i)$  é a distribuição acumulada da função de probabilidade assumida no modelo. No caso da variável resposta  $Y$  ter distribuição discreta torna-se necessário a definição de uma nova variável aleatória,  $u_i$ , com distribuição Uniforme no intervalo  $[u_1, u_2] = [F(y - 1|\theta), F(y|\theta)]$ , em que  $\hat{u}$  é o valor aleatório de uma distribuição Uniforme definido no intervalo  $[\hat{u}_1, \hat{u}_2] = [F(y - 1|\hat{\theta}), F(y|\hat{\theta})]$ . Assim, nos casos em que o modelo ajustado estiver adequado  $\hat{r}_i$  terá distribuição Normal padrão aproximada.

Para analisar a qualidade do ajuste o pacote “gamlss” disponibiliza por meio da função `plot.gamlss()`, que também pode ser acionada com o `plot()`, quatro gráficos que utilizam os resíduos quantílicos aleatorizados, sendo eles dados pelo:

- Resíduos *versus* valores ajustados;
- Resíduos *versus* um índice ou uma variável explicativa;
- Densidade estimada de Kernel (distribuição empírica) dos resíduos;
- Gráfico quantil-quantil normal (QQ-plot) dos resíduos.

Nos casos de ajustes adequados, são esperados que todos esses gráficos apresentem resíduos com distribuição Normal padrão.

O *Worm plot* é outro gráfico disponibilizado pelo pacote “gamlss”, por meio da função `wp()`, que possibilita a realização de um diagnóstico mais geral dos resíduos. Na Figura 8, observa-se um *Worm plot*, no qual o modelo ajustado está adequado. Os pontos representam os resíduos do modelo ajustado, e as curvas elípticas delimitam a região de 95% de confiança. Em modelos com ajustes adequados esperam-se que cerca de 95% destes pontos estejam entre as duas curvas elípticas e no máximo 5% estejam fora. A porcentagem mais elevada de pontos fora das duas curvas elípticas, indica que o modelo ajustado é inadequado para explicar a variável resposta (DONOGHOE; MARSCHNER, 2015).

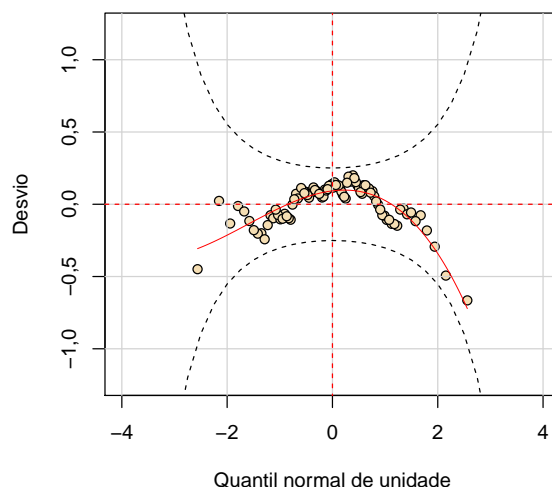


Figura 8 – *Worm plot* de um modelo com ajuste adequado.

Na Tabela 3 são apresentados os tipos de formatos que o *Worm plot* pode assumir em caso de inadequação do ajuste. Além disso, são apresentados também o que cada formato está dizendo com relação aos resíduos e a variável resposta.

Tabela 3 – Descrição dos formatos que o *Worm plot* pode assumir.

Formato <i>Worm plot</i>	Resíduos	Variável resposta
acima da elipse superior	média muito alta	parâmetro de localização muito baixo
abaixo da elipse inferior	média muito baixa	parâmetro de localização muito alto
linha com inclinação positiva	variância muito alta	parâmetro de escala muito baixo
linha com inclinação negativa	variância muito baixa	parâmetro de escala muito alto
U	assimetria positiva	assimetria muito baixa
U invertido	assimetria negativa	assimetria muito alta
S com dobra esquerda para baixo	Leptocúrtica	curtose muito baixa
S com dobra esquerda para cima	Platicúrtica	curtose muito alta

Juntamente com esses gráficos são disponibilizadas, também, as medidas descritivas do resíduo. Para que os resíduos tenha distribuição Normal padrão é esperado que medidas descritivas, tenham a média próxima de zero, com desvio padrão de aproximadamente um e as medidas com relação a assimetria e a curtose sejam próximas de zero e três, respectivamente.

### 2.2.3 Abordagem Bayesiana

A abordagem bayesiana distingue-se da abordagem frequentista por se fundamentar no princípio de que toda incerteza precisa ser modelada por meio de probabilidades, desta forma a

inferência estatística pode tirar conclusões lógicas fundamentando-se nas leis de probabilidade (CHRISTENSEN et al., 2011).

A inferência bayesiana considera não só a amostra mas também o parâmetro  $\theta$  como uma quantidade aleatória, que tem uma distribuição de probabilidades associada. Isso possibilita a combinação de informação *a priori* com a informação proveniente dos dados amostrais por meio do Teorema de Bayes, originando assim a distribuição *a posteriori*, que por sua vez pode se observar o comportamento dos parâmetros após os dados amostrais serem incorporados. Sendo esse comportamento adequado pode-se inferir sobre os resultados dos parâmetros estimados. Utilizando o pacote “INLA” do ambiente estatístico R (R CORE TEAM, 2016) ajustou-se os modelos propostos com a distribuição Binomial e Beta-binomial, as quais foram atribuídas distribuições *a priori* não informativas. A distribuição Multinomial não foi utilizada na abordagem bayesiana devido a não implementação da mesma pelo pacote “INLA”.

### 2.2.3.1 Teorema de Bayes

Toda a construção da teoria bayesiana é consolidada pelo Teorema de Bayes. Para entender esse teorema considera-se  $(\Omega, A, P)$  como um espaço de probabilidade sendo os eventos  $A_1, A_2, \dots, A_n$  uma partição do espaço amostral  $\Omega$ , com  $A_i$  dois a dois disjuntos,  $P(A_i) > 0$  e  $\bigcup_{i=1}^n A_i = \Omega$ ,  $i = 1, 2, \dots, n$ . Dado um outro evento  $B$  qualquer,  $B \subset \Omega$ , e para todo  $j = 1, 2, \dots, n$ , tem-se que o Teorema de Bayes é dado por (DEGROOT et al., 1986):

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)} \quad (2.12)$$

A interpretação do Teorema de Bayes se dá ao considerar que as probabilidades  $P(A_i)$  representam uma informação inicial e subjetiva sobre o evento de interesse  $A$ , a qual será modificada assim que a ocorrência do evento  $B$  for conhecida. Isso mostra que a ocorrência do evento  $B$  modificará as probabilidades *a priori* sobre  $A_i$ , as quais são atualizados pelo Teorema de Bayes, resultando nas probabilidades *a posteriori*  $P(A_i|B)$ .

### 2.2.3.2 Distribuição *a priori* e *a posteriori*

De maneira geral, a distribuição *a priori* é o único fator que diferencia a inferência frequentista da inferência bayesiana. Considera-se então  $Y_1, Y_2, \dots, Y_n$  como uma amostra aleatória independente da variável aleatória  $Y$  que será modelada por meio de uma distribuição de probabilidade com parâmetro  $\theta$  desconhecido, tem-se que toda informação atribuída a esse parâmetro  $\theta$ , antes mesmo da coleta dos dados, é chamada de informação *a priori*, essa informação pode ser traduzida formalmente por uma distribuição de probabilidade *a priori* para  $\theta$ , denotada aqui por  $\pi(\theta)$  (BOX; TIAO, 2011). Essa distribuição *a priori*  $\pi(\theta)$  representa, probabilisticamente, o conhecimento prévio do parâmetro  $\theta$ , antes mesmo de coletar os dados.

Mediante o Teorema de Bayes, a distribuição *a priori* é combinada com a distribuição dos dados amostrados (verossimilhança) e assim surge uma nova distribuição *a posteriori*, que será denotada aqui por  $\pi(\boldsymbol{\theta}|y)$ . Conseqüentemente, toda a inferência é realizada por meio da distribuição *a posteriori* e toda informação pode ser resumida por meio das estimativas pontuais ou intervalares.

Quando o parâmetro  $\boldsymbol{\theta}$  assume valores discretos tem-se que a distribuição *a posteriori* para  $\boldsymbol{\theta}$  dado  $y$ , é dada por:

$$\pi(\boldsymbol{\theta}|y) = \frac{L(\boldsymbol{\theta}|y)\pi(\boldsymbol{\theta})}{\sum_{i \geq 1} L(\boldsymbol{\theta}_i|y)\pi(\boldsymbol{\theta}_i)}, \quad \boldsymbol{\theta} \in \Theta. \quad (2.13)$$

Já quando o parâmetro  $\boldsymbol{\theta}$  assume valores contínuos num dado intervalo, tem-se que a distribuição *a posteriori* de  $\boldsymbol{\theta}$  dado  $y$ , é dada por:

$$\pi(\boldsymbol{\theta}|y) = \frac{L(\boldsymbol{\theta}|y)\pi(\boldsymbol{\theta})}{\int_{\Theta} L(\boldsymbol{\theta}|y)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \quad \boldsymbol{\theta} \in \Theta. \quad (2.14)$$

Tanto o denominador da equação (2.13) como o da equação (2.14) equivale a distribuição marginal da variável aleatória  $Y$ , ou seja, representa a observação  $Y$  independente de quem seja  $\boldsymbol{\theta}$ . Isso equivale dizer que esse denominador do Teorema de Bayes não depende do parâmetro  $\boldsymbol{\theta}$ , sendo portanto uma constante normalizadora. Dessa forma, pode-se reescrever tanto a equação (2.13) como a equação (2.14) da seguinte forma:

$$\pi(\boldsymbol{\theta}|y) \propto L(\boldsymbol{\theta}|y)\pi(\boldsymbol{\theta}) \quad \boldsymbol{\theta} \in \Theta. \quad (2.15)$$

em que  $\propto$  significa “proporcional a”.

### 2.2.3.3 Aproximação de Laplace Aninhada Integrada (INLA)

Proposto por [Rue, Martino e Chopin \(2009\)](#) a Aproximação de Laplace Aninhada Integrada (INLA), do inglês, *Integrated Nested Laplace Approximation* tem a finalidade de proporcionar a realização da inferência bayesiana em campos latentes gaussianos, mais precisamente nos casos em que é assumido independência condicional, em uma subclasse de modelos estruturados aditivos chamados de Campos Aleatórios Markovianos Gaussianos (CAMG), do inglês, *Gaussian Markov Random Fields*.

De acordo com [Rue, Martino e Chopin \(2009\)](#), sendo  $Y_1, Y_2, \dots, Y_n$  uma amostra aleatória independente da variável aleatória  $Y$ , condicionada ao campo latente  $x$ , isto é,

$$p(Y|X, \boldsymbol{\theta}) = \prod_i p(y_i|x_i, \boldsymbol{\theta}) \quad (2.16)$$

em que cada um dos elementos do campo latente tem apenas uma observação contida. No momento que atribui uma distribuição de probabilidade *a priori* para o hiperparâmetro  $\boldsymbol{\theta}$ ,

obtem-se a distribuição *a posteriori*, a qual é dada por:

$$p(X, \boldsymbol{\theta}|Y) \propto p(\boldsymbol{\theta})p(X|\boldsymbol{\theta}) \prod_i p(y_i|x_i, \boldsymbol{\theta}) \quad (2.17)$$

$$p(X, \boldsymbol{\theta}|Y) \propto p(\boldsymbol{\theta})|Q(\boldsymbol{\theta})|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} X' Q(\boldsymbol{\theta}) X + \sum_i \log(p(y_i|x_i, \boldsymbol{\theta})) \right\}, \quad (2.18)$$

sendo  $Q(\boldsymbol{\theta})$  a matriz de precisão.

Na abordagem INLA a aproximação das distribuições *a posteriores* marginais para as quantidades latentes  $p(x_i|Y)$ ,  $\forall_i = 1, \dots, n$  e para os hiperparâmetros do modelo gaussiano latente  $p(\theta_j|Y)$ ,  $\forall_j = 1, \dots, m$ , é dado por:

$$\pi(\boldsymbol{\theta}|y) = p(x_i|Y) = \int p(\boldsymbol{\theta}|Y)p(x_i|\boldsymbol{\theta}, Y)d\boldsymbol{\theta} \quad (2.19)$$

$$p(\theta_j|Y) = \int p(\boldsymbol{\theta}|Y)d\theta_{-j} \quad (2.20)$$

Dado que as equações (2.19) (2.20) não possui solução analítica, tem-se que a sua solução será obtida pela aproximação de Laplace.

A metodologia INLA fornece uma alternativa rápida e determinística com relação a Cadeia de Markov Monte Carlo (MCMC) (SORENSEN; GIANOLA, 2007) que, até então, é uma ferramenta padrão na inferência bayesiana. A principal vantagem da aproximação via INLA sobre MCMC é a sua eficiência e rapidez com relação aos resultados obtidos, pois enquanto o INLA retorna respostas em poucos instantes o MCMC pode exigir horas ou até dias. Além disso, por ser um método determinístico e baseado em amostragem não tem a necessidade de verificação de convergência.

A descrição completa da metodologia INLA está disponível em Rue, Martino e Chopin (2009).

#### 2.2.3.3.1 Seleção e validação de modelo

A identificação do modelo que melhor explicam os dados observados, constitui tanto na abordagem frequentista quanto na abordagem bayesiana uma parte importantíssima da modelagem estatística, dando ao pesquisador indícios de qual dos modelos ajustados melhor se adequou ao seu conjunto de dados. Essa comparação e validação dos modelos estimados por inferência bayesiana será realizada com o auxílio do Critério de Informação da Deviance (DIC), Critério de Informação Watanabe-Akaike (WAIC), Pseudo Marginal Log Verossimilhança (LPML), Distribuição Condicional Preditiva Ordenada (CPO) e a Transformação Integral de Probabilidade (PIT):

- DIC: Critério foi proposto por Spiegelhalter et al. (2002) com a finalidade de comparar modelos bayesianos, sendo essa comparação embasada pela distribuição *a posteriori* da

deviance de cada modelo. E é definido como:

$$DIC = \bar{D} + 2pD,$$

em que  $\bar{D}$  é a média *a posterior* do desvio do modelo e  $pD$  é a penalização do modelo (número efetivo de parâmetros), dada pela diferença entre a média *a posteriori* da deviance e a deviance das médias *a posteriori* dos parâmetros do modelo de interesse. Sendo este um dos critérios de ajuste mais utilizados dentro da estatística bayesiana, tem-se que o melhor ajuste do modelo aos dados será dado pelo que apresentar o menor valor de DIC;

- WAIC: Proposto por [Watanabe \(2010\)](#), este é um método totalmente bayesiano, o qual aproxima-se da validação cruzada bayesiana, além disso é invariante para parametrização e funciona também para modelos singulares. Considere-se  $y_1, y_2, \dots, y_n$  como uma amostra aleatória, modelada por uma distribuição de probabilidade com parâmetro  $\theta$ . Assim,  $p(\mathbf{y}|\theta) = \prod_{i=1}^n p(y_i|\theta)$ , e suponha também uma distribuição *a priori*  $\pi(\theta)$ , logo a distribuição *a posteriori* é dada por  $\pi(\theta|y) = \prod_{i=1}^n p(y_i|\theta)\pi(\theta)$  e a distribuição preditiva *a posterior* é definida por  $\pi(\hat{y}) = p(\hat{y}|\theta)\pi(\theta)$ . Para manter a comparabilidade com o conjunto de dados e para obtê uma interpretação mais fácil das diferenças na escala do número efetivo de parâmetros, define-se uma medida conhecida por esperança do logaritmo da precisão preditiva (ELPP) para os  $n$  pontos amostrados, dada por:

$$ELPP = \sum_{i=1}^n E_{f_i}(\log\pi(\hat{y}_i)), \quad (2.21)$$

em que  $f(y_i)$  representa a função de distribuição dos dados. E essa distribuição depende de  $\theta$  desconhecido, o qual será utilizado o WAIC ou validação cruzada para aproximar (2.21). O logaritmo da precisão preditiva é definido por:

$$LPP = \sum_{i=1}^n \log \int p(y_i|\theta)\pi(\theta)d\theta \quad (2.22)$$

O LPP dos dados observados é superestimada pela ELPP para dados futuros. Assim, na prática calcula-se a esperança de  $\pi(\theta)$  utilizando simulação, que será denotado aqui por  $\theta^s$  para  $s = 1, 2, \dots, S$ , sendo  $S$  a quantidade de simulação. Assim, a LPP estimada é dada por:

$$\widehat{LPP} = \sum_{i=1}^n \log \left( \frac{1}{S} \sum_{s=1}^S p(y_i|\theta^s) \right) \quad (2.23)$$

E finalmente o WAIC é definido da seguinte forma:

$$WAIC = -2(\widehat{LPP} - k), \quad (2.24)$$

em que  $k$  é o número efetivo de parâmetros estimados. Sendo a interpretação entendida como o modelo que apresentar o menor valor de WAIC será indicado como o modelo que melhor se ajustou aos dados (VEHTARI; GELMAN, 2014);

- CPO: A distribuição Condicional Preditiva ordenada, do inglês *Conditional Predictive Ordinate Distribution* (CPO), pode ser escrita da seguinte forma:

$$CPO_i = p(y_i|y_{-i}) = \int f(y_i|\theta)f(\theta|y_{-i})d\theta, \quad (2.25)$$

em que a densidade preditiva representa uma medida do quão provável seria obter uma futura observação igual a  $y_i$  dado a amostra  $y_{-i}$ , sendo  $y_{-i}$  o conjunto de dados obtido ao excluir-se  $y_i$ , para  $i = 1, 2, \dots, n$ . Como para muitos modelos estatísticos a medida  $CPO_i$  não tem uma forma analítica fechada Gelfand e Dey (1994) apresentaram uma forma de aproximar essa quantidade, por meio de uma amostra *a posteriori* de  $\theta$ , a qual é dada por:

$$\widehat{CPO}_i = \left( \frac{1}{n} \sum_{j=1}^n \frac{1}{f(y_i|\theta^{(j)})} \right)^{-1}, \quad (2.26)$$

em que  $n$  é o tamanho da amostra e  $\theta^{(j)}$  é o  $j$ -ésimo valor obtido da distribuição *a posteriori* de  $\theta$ ;

- PIT: Foi proposto por Rosenblatt (1952) como uma medida de qualidade de ajuste. E o seu calculo é dado por:

$$\widehat{PIT}_i = p(y_i^{new} \leq y_i|y_{-i}) \quad (2.27)$$

A avaliação do desempenho preditivo do modelo pode ser feita por meio da distribuição empírica da PIT. A distribuição preditiva do modelo ajustado será considerada coerente se a distribuição empírica da PIT se aproximar da Uniforme(0,1), (GNEITING; RAFTERY, 2007);

- LPML: Esse é outro critério de comparação útil na modelagem bayesiana. E o seu calculo é realizado utilizando a medida CPO, sendo definida da seguinte forma:

$$LPML = \sum_{i=1}^n \log(\widehat{CPO}_i). \quad (2.28)$$

A interpretação deste critério é de que quanto maior o valor de LMPL, melhor é a adequação do modelo proposto ao conjunto de dados.

---

**CAPÍTULO 3**

---

---

**RESULTADOS E DISCUSSÕES**

---

**3.1 Análise exploratória**

Por meio da análise descritiva busca-se entender o comportamento das plântulas normais, anormais e sementes não germinadas com relação a cada um dos possíveis fatores de influência individualmente. Desta forma, observou-se que a proporção de sementes que apresentaram plântulas normais foi de 55,69%. As plântulas anormais 12,72%, sendo esse registro inferior a proporção de sementes não germinadas que foi de 31,59%, conforme resultados apresentados na Figura 9.

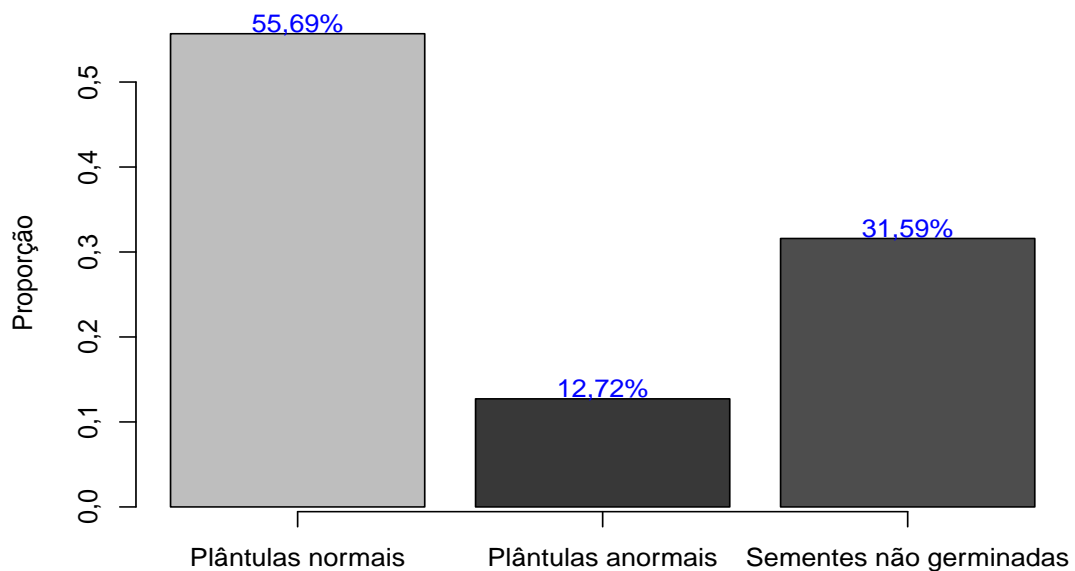


Figura 9 – Proporção do tipo de plântulas normais, anormais e sementes não germinadas observadas no experimento.



A Figura 10 permite analisar o desempenho da germinação com relação a cada um dos possíveis fatores de influência descritos na Tabela 1. Analisando individualmente, tem-se que o germinador 3 e 4 parece estar produzindo mais plântulas normais que o germinador 1 e 2. Já com relação aos lotes também é possível observar uma diferença, uma vez que os lotes 2 e 1 aparentam estar proporcionando uma maior quantidade de plântulas normais do que os lotes 3 e 4. Quando analisa a posição e a luz individualmente, com relação a quantidade de plântulas normais, graficamente não é possível verificar diferença.

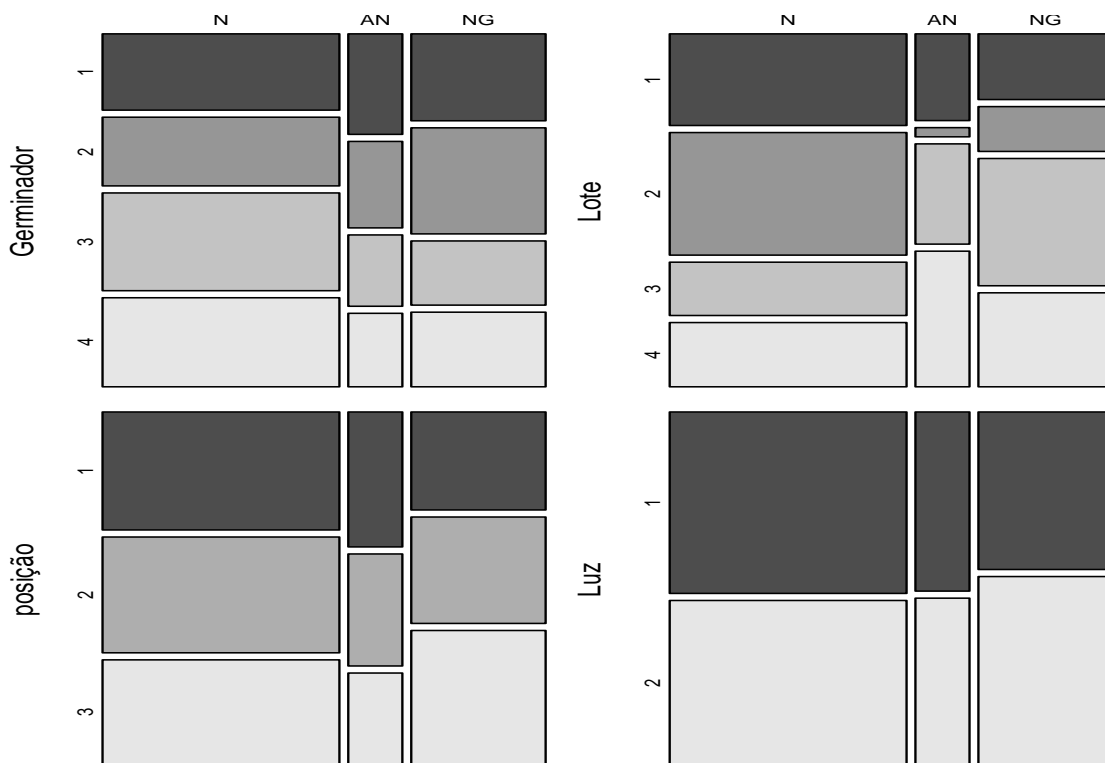


Figura 10 – Comportamento do tipo de plântulas normais (N), anormais (AN) e sementes não germinadas (NG) dentro do experimento levando em consideração os fatores como germinador, luz, posição e lote.

### 3.2 Resultados

Com o objetivo de entender o tipo de plântulas produzidas pelas sementes da *stevia rebaudiana* com relação aos fatores de influência foram ajustados modelos de regressão com efeitos mistos, atribuindo efeito aleatório as variáveis lote e germinador. Para isso, ajustou a variável resposta (Tipo de plântulas produzidas) em relação aos possíveis fatores de influência utilizando as distribuições Binomial, Beta-Binomial e Multinomial com abordagem frequentista e as distribuições Binomial e Beta-binomial com abordagem bayesiana, com distribuições *a priori* não informativas. Para os ajustes com as distribuições Binomial e Beta-Binomial, agrupou-se os níveis da variável resposta (Tipo de plântulas produzidas), no entanto para a regressão Multinomial manteve-se os níveis originais, conforme a descrição apresentada na Figura 11.

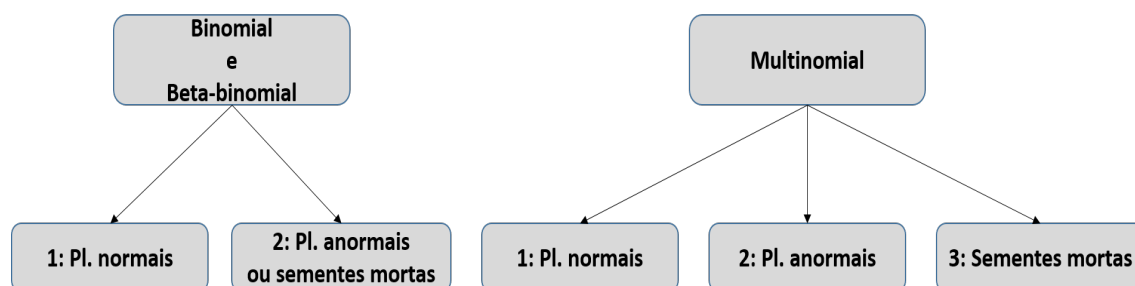


Figura 11 – Categorização utilizada do tipo de plântulas produzidas para cada uma das distribuições utilizadas.

Para cada um das três distribuições adotadas foram ajustados seis modelos de regressão, os quais estão descritos na Tabela 4.

Tabela 4 – Modelos de regressão a serem ajustados considerando as distribuições Binomial, Beta-binomial e Multinomial.

ID	Modelos
$M_0$	$\eta = \beta_0$
$M_1$	$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3$
$M_2$	$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2$
$M_3$	$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Lote$
$M_4$	$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Germinador$
$M_5$	$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Lote + b_2 Germinador$

### 3.2.1 Regressão Binomial

Considerando o modelo de regressão Binomial, ajustou-se seis modelos, Tabela 4, com a finalidade de identificar a relação existente entre a plântulas normais das sementes de *stevia rebaudiana* e os seus possíveis fatores de influência como posição, luz, lote e germinador. A seleção do modelo por meio da análise frequentista foi baseada nas medidas referentes ao Critério de Informação de Akaike (AIC), Critério de Informação de Akaike Corrigido (AIC<sub>c</sub>), Deviance Global (DG) e no Critério bayesiano de Schwarz (BIC). Enquanto, na análise bayesiana baseou-se no Critério de Informação Deviance (DIC), pelo Critério de Informação Watanabe-Akaike (WAIC) e pela pseudo marginal log verossimilhança (LPML). Os valores de todas essas medidas para cada um dos modelos propostos estão apresentados na Tabela 5.

Tabela 5 – Critérios frequentistas e bayesianos para comparação e seleção dos modelos ajustados com a distribuição Binomial.

Modelo	Critérios frequentistas				Critérios bayesianos		
	AIC	AIC <sub>c</sub>	DG	BIC	DIC	WAIC	LPML
$M_0$	13260,1	13260,1	13258,1	13267,3	13260,13	13260,13	-6630,07
$M_1$	13236,6	13236,6	13230,6	13258,2	13236,63	13236,64	-6618,32
$M_2$	13213,6	13213,6	13205,6	13242,3	13213,60	13213,60	-6606,80
$M_3$	11917,8	11917,8	11901,8	11975,1	11915,84	11915,85	-5957,92
$M_4$	13034,9	13034,9	13019,0	13092,0	13031,02	13031,02	-6515,51
$M_5$	11925,7	11925,8	11901,8	12011,4	11917,26	11917,27	-5958,63

Conforme os resultados apresentados pela Tabela 5 pode-se perceber que o modelo  $M_3$  foi o que apresentou o menor valor de AIC, AIC<sub>c</sub>, DG e BIC entre os seis modelos ajustados, sendo portanto o modelo que apresentou o melhor ajuste com a regressão Binomial, referente a abordagem frequentista. No ponto de vista bayesiano o modelo  $M_3$  também apresentou-se como o modelo que melhor-se ajustou aos dados, isso se deve ao menor valor de DIC e o menor de WAIC entre os seis modelos ajustados. Além disso, observa-se também que o modelo  $M_3$  possui a maior LPML, isso denota que esse modelo tem maior capacidade preditiva, quando comparado com os outros modelos.

Na Tabela 6, pode-se observar as estimativas frequentistas dos parâmetros do modelo  $M_3$ , selecionado, com o seus respectivos intervalos de confiança, como também as estimativas bayesianas com os seus respectivos desvios padrão e intervalos de credibilidade. Ambas as abordagens retornaram estimativas bem parecidas, mais vale destacar que a abordagem frequentista retornou intervalos de confiança mais preciso que os intervalos de credibilidade da abordagem bayesiana. Do ponto de vista frequentista apenas o parâmetro  $\beta_1$  não foi significativo, já do ponto de vista bayesiano observa-se que os parâmetros  $\beta_0$  e  $\beta_1$  não significativos.

Tabela 6 – Estimativas pontuais e intervalares do ponto de vista frequentista e bayesiano para os coeficientes de regressão utilizados para ajustar o parâmetro  $\pi$  da distribuição  $BI(n, \pi)$ .

Variável associada	Parâmetro	Estimativas frequentista		Estimativas bayesiana		
		Estimativa	IC (95%) Perfilado	Média <i>a posteriori</i>	Desvio Padrão	IC <sub>r</sub> (95%)
-	$\beta_0$	0,49	( 0,4133; 0,5924)	0,49	0,72	(-0,9869; 1,9635)
Posição 2	$\beta_1$	-0,01	(-0,1300; 0,0992)	-0,01	0,05	(-0,1147; 0,0971)
Posição 3	$\beta_2$	-0,27	(-0,3801; -0,1814)	-0,27	0,05	(-0,3724; -0,1612)
Luz 2	$\beta_3$	-0,23	(-0,3222; -0,1602)	-0,23	0,04	(-0,3171; -0,1443)
Lote	$\sigma_b^2$	0,93	-	1,11	0,81	(0,1109; 3,1700)

A validade deste modelo ajustado tanto com a abordagem frequentista quanto com a abordagem bayesiana só pode ser confirmada após a verificação dos pressupostos do mesmo,

desta forma será apresentada a análise dos resíduos para o modelo  $M_3$  referente aos dois tipos de abordagem utilizada.

Na abordagem frequentista, a análise baseou-se nos resíduos quantílicos aleatorizados e nos gráficos de diagnóstico que estão sendo apresentados na Figura 12. O gráfico dos valores ajustados pelos quantílicos residuais está mostrando que os resíduos estão distribuídos aleatoriamente em torno de zero. O gráfico da densidade estimada e o gráfico Q-Q plot estão indicando que os quantis residuais têm distribuição Normal padrão, isso atende o esperado para um ajuste adequado.

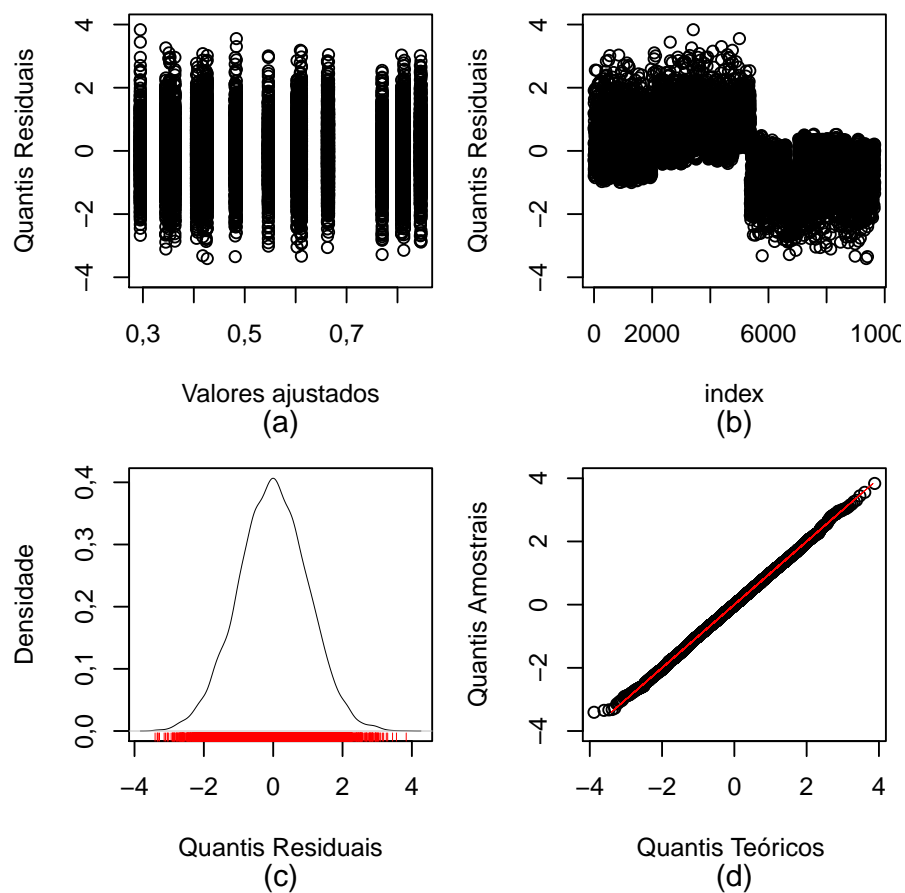


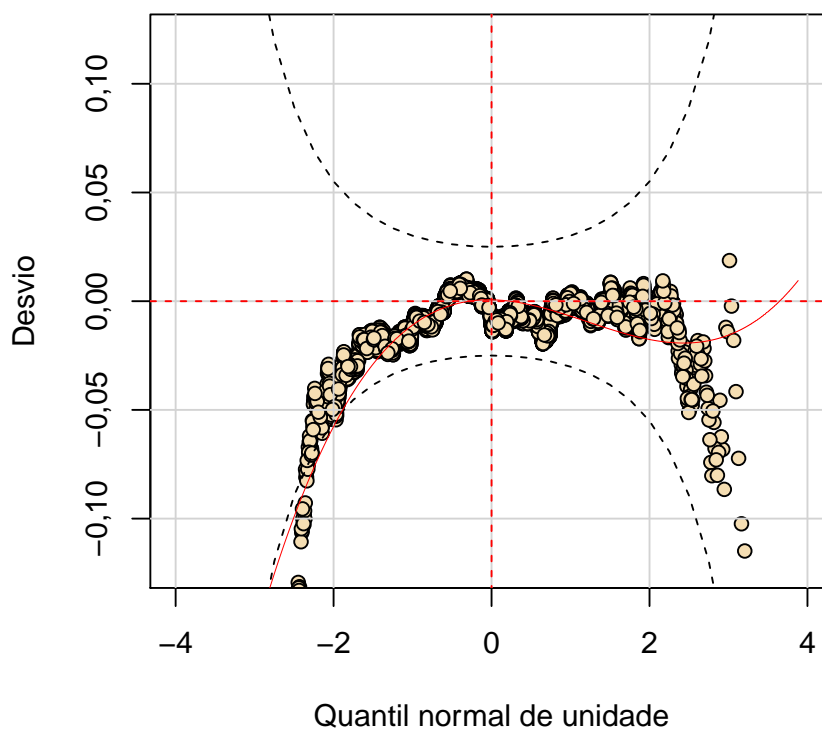
Figura 12 – Gráficos para diagnóstico do ajuste do modelo considerando a distribuição Binomial para os dados.

De acordo com o resumo apresentado na Tabela 7, as medidas descritivas do resíduo quantílico aleatorizado, com valor da média próxima de 0, a variância próxima de 1, bem como, os coeficientes de assimetria e curtose de aproximadamente 0 e 3, respectivamente, estão todas dentro do esperado para uma variável aleatória com distribuição Normal padrão.

Tabela 7 – Medidas descritivas do resíduo quantílico aleatorizado.

Média	=	-0,00892
Variância	=	1,01506
Coef. de assimetria	=	-0,05826
Coef. de curtose	=	3,07060

O *Worm plot*, apresentado na Figura 13, do resíduo quantílico aleatorizado referente a regressão Binomial só vem reforçar a adequabilidade do modelo ajustado.

Figura 13 – *Worm plot* do resíduo quantílico aleatorizado referente a regressão Binomial.

A verificação da qualidade do ajuste referente a abordagem bayesiana, foi realizada por meio do gráfico dos valores da medida preditiva PIT, bem como o seu histograma, conforme-se observa na Figura 14. Apesar dos resultados Bayesianos estarem bem próximos dos resultados Frequentistas, verifica-se que a distribuição dos valores de PIT não estão dentro do esperado de um modelo bem ajustado, conforme pode-se observar na Figura 14 (a) e (b). Os valores de PIT não tem distribuição Uniforme (0,1).

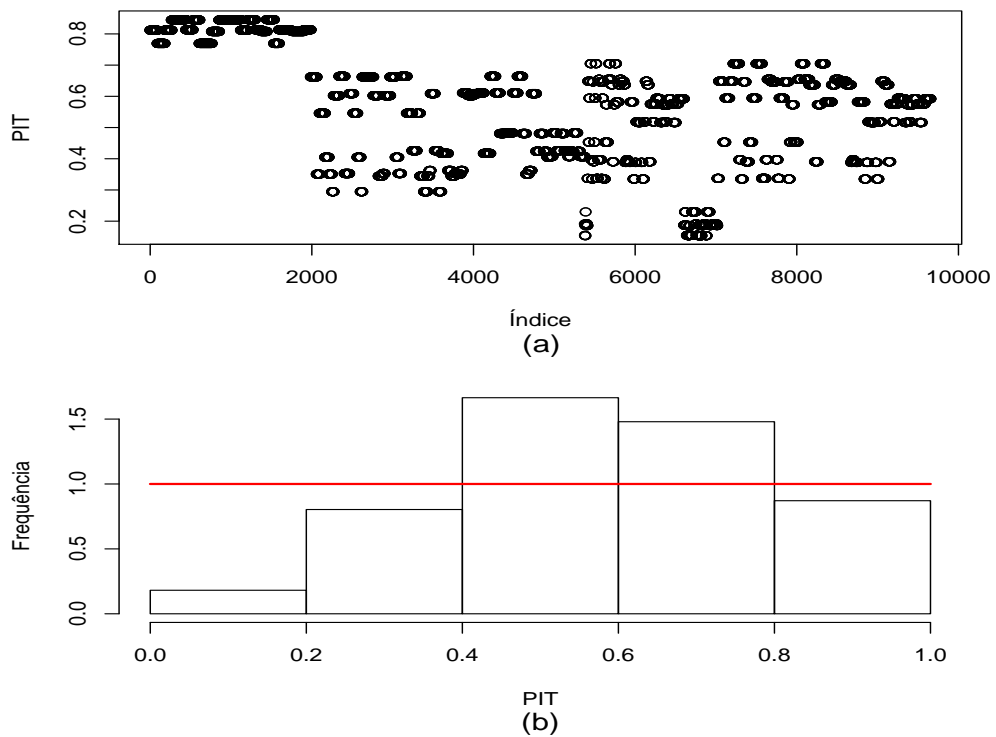


Figura 14 – (a) Gráfico dos valores de PIT para o modelo Binomial. (b) Histograma dos valores de PIT para o modelo Binomial.

A Figura 15 apresenta as densidades marginais *a posteriori* referente aos parâmetros  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  e  $\beta_3$ . O comportamento das densidades para todos os parâmetros é considerado suave, ou seja, contém características desejáveis sem presença de bimodalidade e todas simétricas.

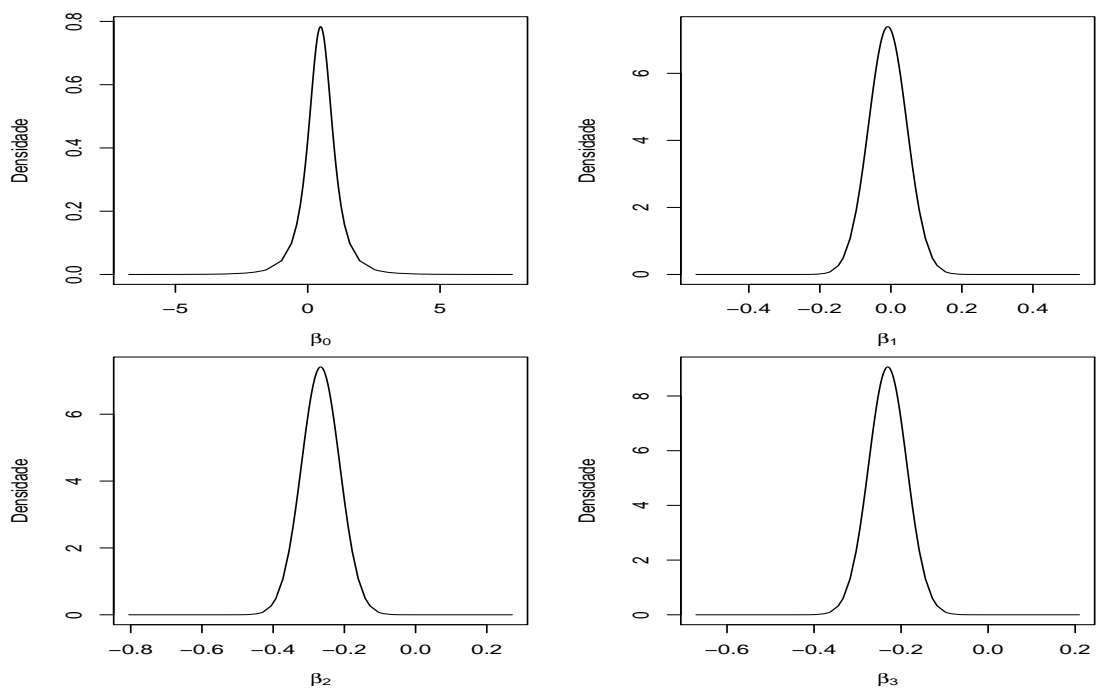


Figura 15 – Densidades marginais *a posteriori* para  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  e  $\beta_3$ .

### 3.2.2 Regressão Beta-binomial

Foram ajustados seis modelos, Tabela 4, com a distribuição Beta-binomial, com isso buscou-se identificar a relação existente entre a plântulas normais das sementes de *stevia rebaudiana* e os possíveis fatores que podem influenciar a plântulas como posição, luz, lote e germinador. Para escolher um modelo mais parcimonioso utilizou-se as medidas AIC,  $AIC_c$ , DG e BIC na abordagem frequentista e as medidas DIC, WAIC e LPML na abordagem bayesiana, em que todos os valores destas medidas, para cada um dos modelos ajustados, estão descritos na Tabela 8.

Tabela 8 – Medidas referente a abordagem frequentista e bayesiana na comparação e seleção dos seis modelos ajustados com a distribuição Beta-binomial.

Modelo	Critérios frequentistas				Critérios bayesianos		
	AIC	$AIC_c$	DG	BIC	DIC	WAIC	LPML
$M_0$	13262,1	13262,14	13258,1	13276,4	13260,13	13260,13	-6630,07
$M_1$	13238,6	13238,6	13230,6	13267,3	13236,63	13236,64	-6618,32
$M_2$	13215,6	13215,6	13205,6	13251,5	13213,60	13213,60	-6606,80
$M_3$	11919,8	11919,8	11901,8	11984,4	11915,84	11915,85	-5957,92
$M_4$	13036,9	13036,9	13019,0	13101,2	13031,02	13031,02	-6515,51
$M_5$	11927,8	11927,8	11901,8	12020,9	11917,09	11917,10	-5958,55

As estimativas do modelo  $M_3$ , o mais parcimonioso, com os seus respectivos intervalos de confiança de 95% da abordagem frequentista e as estimativas, desvios padrão e intervalos de credibilidade da abordagem bayesiana estão apresentados na Tabela 9. A função de ligação *logit* foi utilizada para ligar o parâmetro de posição  $\pi$  ao preditor linear  $\eta$ . Observa-se que os intervalos de confiança frequentista são mais precisos que os intervalos de credibilidade da abordagem bayesiana. Do ponto de vista frequentista apenas o parâmetro  $\beta_1$  não foi significativo, já com relação ao enfoque bayesiano obteve-se dois parâmetros não significativos,  $\beta_0$  e  $\beta_1$ .

Tabela 9 – Estimativas pontuais e intervalares do ponto de vista frequentista e bayesiano para os coeficientes de regressão utilizados para ajustar o parâmetro  $\pi$  da distribuição  $BB(\pi, \sigma)$ .

Variável associada	Parâmetro	Estimativas frequentistas		Estimativas bayesianas		
		Estimativa	IC (95%) Perfilado	Média <i>a posteriori</i>	Desvio Padrão	IC <sub>r</sub> (95%)
-	$\beta_0$	0,62	(0,5379; 0,7146)	0,49	0,72	(-0,9695; 1,9459)
Posição 2	$\beta_1$	-0,01	(-0,1010; 0,0808)	-0,01	0,05	(-0,1147; 0,0971)
Posição 3	$\beta_2$	-0,27	(-0,3641; -0,1412)	-0,27	0,05	(-0,3724; -0,1612)
Luz 2	$\beta_3$	-0,23	(-0,2998; -0,1366)	-0,23	0,04	(-0,3171; -0,1443)
Lote	$\sigma_b^2$	0,93	-	1,11	0,9110	(0,1358; 3,4867)

Na Tabela 10 tem-se a estimativa do modelo de regressão com relação ao parâmetro de escala  $\sigma$ , em que a função de ligação *log* foi utilizada para ligar o parâmetro de escala  $\sigma$  ao preditor linear  $\eta$ .

Tabela 10 – Estimativas pontuais e intervalares do ponto de vista frequentista e bayesiano para os coeficientes de regressão utilizados para ajustar o parâmetro  $\sigma$  da distribuição  $BB(\pi, \sigma)$ .

Variável associada	Parâmetro	Estimativas frequentistas		Estimativas bayesiano		
		Estimativa	IC (95%) Perfilado	Média <i>a posteriori</i>	Desvio Padrão	IC <sub>r</sub> (95%)
-	$\beta_0$	-0,10	(-0.8833; 0.7002)	0,50	0,27	(0,0432;0,9573)

Para validar o modelo ajustado será apresentado agora a análise dos resíduos, tanto para a abordagem frequentista quanto para a bayesiana.

Na frequentista, a análise foi baseada nos resíduos quantílicos aleatorizados e os gráficos de diagnóstico são dados pela Figura 16. O gráfico dos valores ajustados pelos resíduos estão dentro do esperado para uma variável aleatória com distribuição Normal, pois todos os pontos estão distribuídos aleatoriamente em torno de zero. Tanto o gráfico da densidade estimada como o gráfico Q-Q plot indicam que os quantis residuais têm distribuição Normal padrão.

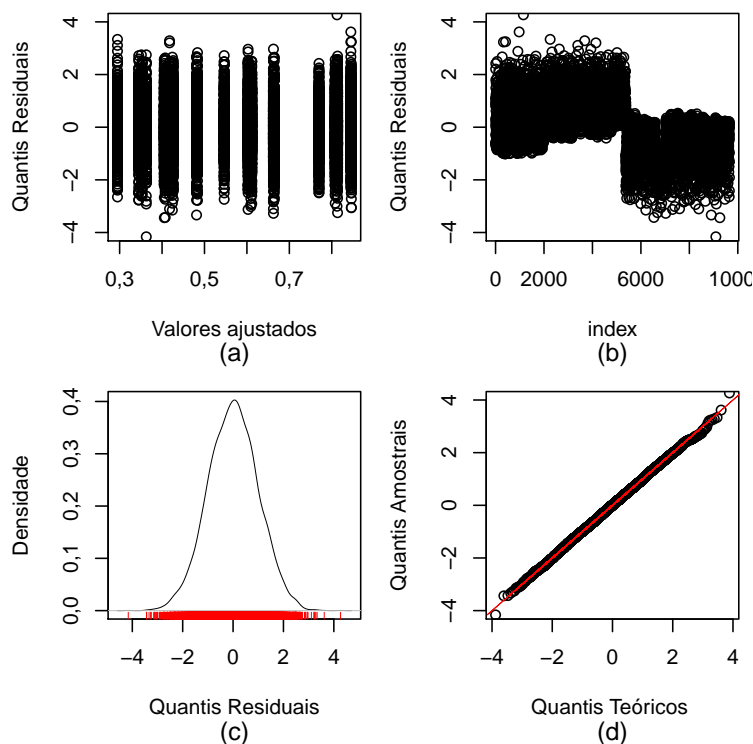


Figura 16 – Gráficos para diagnóstico do ajuste do modelo considerando a distribuição Beta-binomial para os dados.



As medidas descritivas estão confirmando as informações obtidas dos resíduos quantílicos aleatorizados (Figura 16), pois tanto a média, quanto a variância e os coeficientes de curtose e de correlação são valores esperados para uma variável aleatória com distribuição Normal padrão, conforme se observa na Tabela 11 .

Tabela 11 – Resumo do resíduo quantílico aleatorizado.

Média	=	-0,00557
Variância	=	1,00051
Coef. de assimetria	=	-0,00413
Coef. de curtose	=	3,07755

Conforme *Worm plot*, apresentado na Figura 17, o modelo de regressão Beta-binomial é adequado na descrição da germinação de sementes com relação aos seus fatores de influência, pois os pontos dos resíduos estão distribuídos dentro da faixa que indica adequabilidade do modelo.

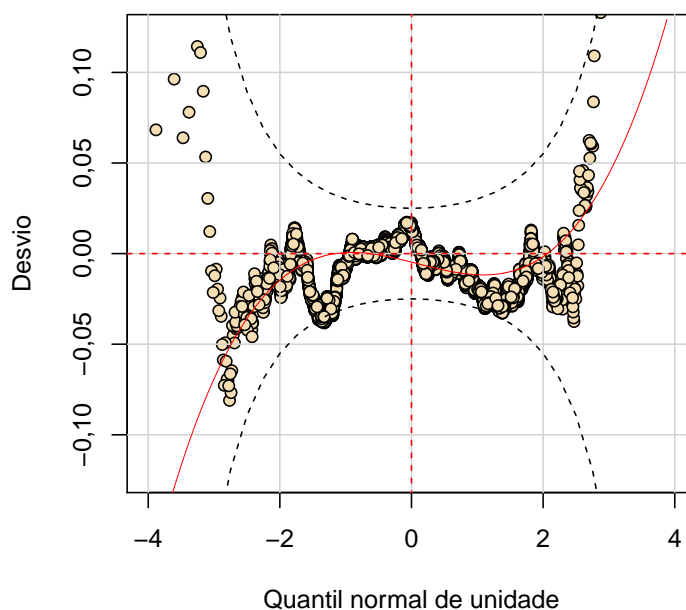


Figura 17 – *Worm plot* do ajuste com a distribuição Beta-binomial.

Na abordagem bayesiana a validação do modelo foi realizada por meio do gráfico dos valores da medida preditiva PIT e pelo seu histograma, conforme-se observa na Figura 18. Mesmo que os resultados Bayesianos estejam bem próximos dos resultados Frequentistas, verifica-se que a distribuição dos valores de PIT não estão dentro do esperado de um modelo bem ajustado, conforme pode-se observar na Figura 18 (a) e (b). Os valores de PIT não apresentaram distribuição Uniforme (0,1).

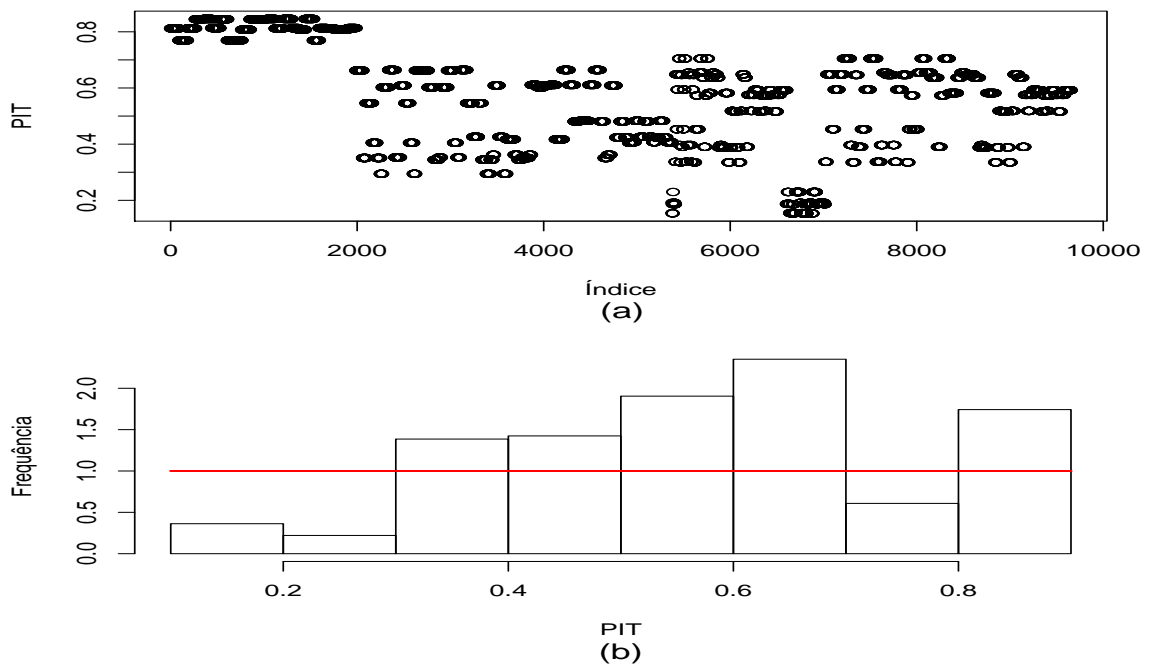


Figura 18 – (a) Gráfico dos valores de PIT para o modelo Beta-binomial. (b) Histograma dos valores de PIT para o modelo Beta-binomial.

As densidades marginais *a posteriori* referente aos parâmetros  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  e  $\beta_3$ , são apresentadas na Figura 19. As densidades para todos os parâmetros revela-se um comportamento suave, sendo essa uma característica desejável.

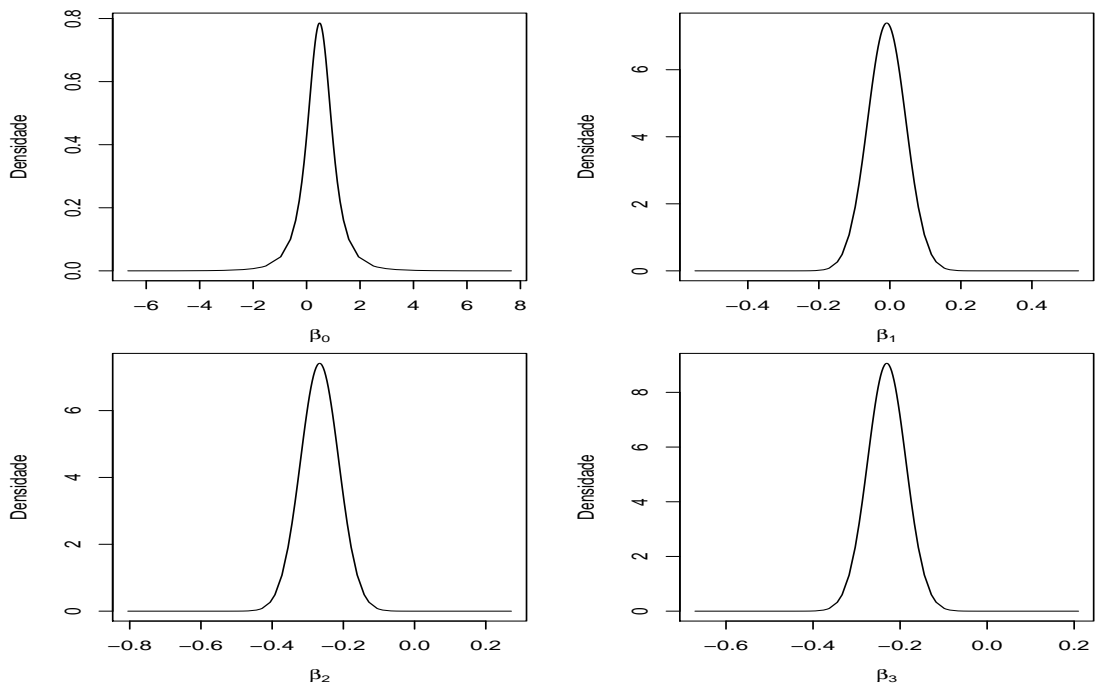


Figura 19 – Densidades marginais *a posteriori* para  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  e  $\beta_3$ .

### 3.2.3 Regressão Multinomial

Atribuindo a distribuição Multinomial a variável resposta tipo de germinação ajustou-se seis modelos de regressão do ponto de vista frequentista, com o objetivo de identificar a relação existente entre a plântulas normais, anormais e sementes não germinadas, bem como os possíveis fatores que podem influenciar esse tipo de plântulas. A distribuição Multinomial neste caso contém três parâmetros, sendo eles  $\pi_1$  proporção de plântulas normais,  $\pi_2$  proporção de plântulas anormais e  $\pi_3$  proporção de sementes não germinadas. A escolha do modelo que melhor se ajustou aos dados foi realizada pelas medidas de AIC,  $AIC_c$ , DG e BIC, esses valores estão descritos na Tabela 12.

Como pode-se observar o modelo  $M_3$  foi o que apresentou as menores medidas em todos os critérios utilizados, sendo esse o modelo final escolhido.

Tabela 12 – Medidas para comparação e seleção dos modelos ajustados com a distribuição Multinomial.

ID	Modelos	AIC	$AIC_c$	DG	BIC
$M_0$	$\eta = \beta_0$	18391,36	18391,36	18387,36	18405,71
$M_1$	$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3$	18299,50	18299,51	18287,50	18342,55
$M_2$	$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2$	18264,56	18264,58	18248,56	18321,96
$M_3$	$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Lote$	16780,30	16780,36	16748,46	16894,55
$M_4$	$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Ger$	18085,86	18085,89	18062,03	18171,34
$M_5$	$\eta = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Lote + b_2 Ger$	16788,18	16788,26	16748,46	16930,65

Observa-se na Tabela 13 as estimativas e os respectivos intervalos de confiança dos parâmetros do modelo selecionado para as plântulas normais em comparação as sementes não germinadas e a proporção de plântulas anormais com relação as sementes não germinadas. A função de ligação utilizada neste caso foi a *logit*.

Tabela 13 – Estimativas pontuais e intervalares do modelo de regressão Multinomial referente a proporção de plântulas normais com relação a sementes não germinadas e da proporção de plântulas anormais com relação a sementes não germinadas.

	Variável associada	Parâmetro	Estimativa	Intervalo de confiança 95% Perfilado
$\frac{\pi_1}{\pi_3}$	-	$\beta_0$	1,22	(1,1393; 1,3123)
	Posicao2	$\beta_1$	-0,09	(-0,2017; 0,0101)
	Posicao3	$\beta_2$	-0,46	(-0,5652; -0,3545)
	Luz2	$\beta_3$	0,30	(-0,3884; -0,2159)
	Lote	$\sigma_b^2$	0,72	-
$\frac{\pi_2}{\pi_3}$	-	$\beta_0$	-0,50	(-0,6141; -0,3826)
	Posicao2	$\beta_1$	-0,27	(-0,4196; -0,1288)
	Posicao3	$\beta_2$	-0,70	(-0,8581; -0,5538)
	Luz2	$\beta_3$	0,25	(-0,3785; -0,1327)
	Lote	$\sigma_b^2$	0,76	-

O modelo ajustado só é válido após a verificação dos pressupostos dos resíduos. A análise dos resíduos deste ajuste foi baseado nos resíduos quantílicos aleatorizados e os gráficos para a verificação dos mesmos estão apresentados na Figura 20. O gráfico dos valores ajustados pelos quantílicos residuais está dentro do esperado que são valores distribuídos aleatoriamente em torno de zero. Os gráficos da densidade estimada e o Q-Q plot também estão dentro do esperado, porque ambos estão indicando que os quantis residuais apresentam ter distribuição Normal padrão.

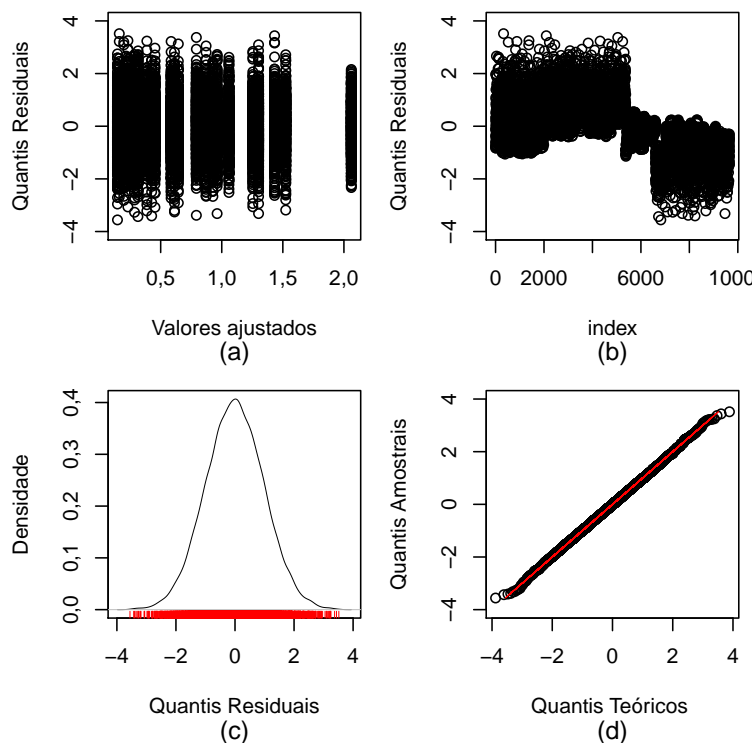


Figura 20 – Gráficos para diagnóstico do ajuste do modelo considerando a distribuição Multinomial para os dados.

As medidas descritivas dos resíduos quantílicos aleatorizados indicam adequação do modelo, porque tanto a média, como a variância e os coeficientes de curtose e de correlação estão com valores dentro do esperados para uma variável aleatória com distribuição Normal padrão, conforme se observa na Tabela 14 .

Tabela 14 – Resumo do resíduo quantílico aleatorizado.

Média	=	0,00205
Variância	=	1,00165
Coef. de assimetria	=	0,03263
Coef. de curtose	=	3,03242

Apresenta-se na Figura 21 o gráfico *Worm plot* fazendo assim um resumo geral dos resíduos quantílicos aleatorizados, no qual o modelo de regressão Multinomial se mostra ade-

quado para descrever a germinação das sementes de *stevia rebaudiana* com relação aos seus fatores de influência.

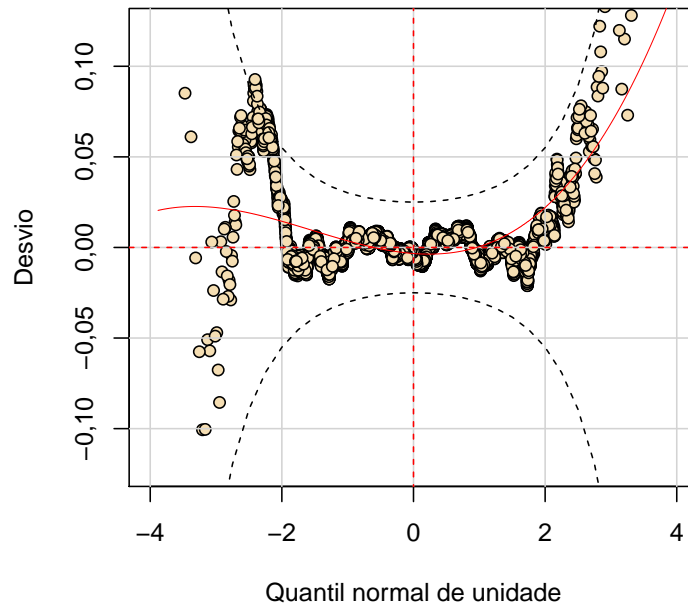


Figura 21 – *Worm plot* do ajuste com a distribuição Multinomial.

### 3.2.4 Discussões

Dado o ajuste, seleção e verificação da adequabilidade dos modelos de regressão Binomial, Beta-binomial e Multinomial na descrição do comportamento da plântulas normais, anormais e sementes não germinadas com relação aos seus fatores de influência, agora pode-se inferir por meio dos resultados obtidos. Pelos resultados apresentados na Tabela 15 pode-se observar que a distribuição Binomial e Beta-binomial não apresentaram diferença significativa, ou seja, ambas estão descrevendo de forma equivalente o comportamento das observações da variável resposta tipo de germinação, a qual é composta neste caso por dois níveis. O modelo Multinomial também está descrevendo de forma adequada a variável tipo de germinação, porém esse modelo está levando em consideração os três níveis do tipo de germinação.

Tabela 15 – Medidas de seleção de modelo referente ao modelo  $M_3$  ajustado com as distribuições Binomial, Beta-binomial e Multinomial.

Modelo	Critérios frequentistas				Critérios bayesianos		
	AIC	AIC <sub>c</sub>	DG	BIC	DIC	WAIC	LPML
Binomial	11917,8	11917,8	11901,8	11975,1	11915,84	11915,85	-5957,92
Beta-binomial	11919,8	11919,8	11901,8	11984,4	11915,84	11915,85	-5957,92
Multinomial	16780,3	16780,4	16748,4	16894,5	-	-	-

A seguir, uma interpretação das estimativas obtidas considerando os modelos de regressão Binomial, Beta-binomial e Multinomial é dada. O modelo de regressão Binomial e Beta-binomial são equivalentes estatisticamente, desta forma a interpretação será uma só para ambas.

O ajuste do modelo de regressão Binomial e Beta-binomial utilizou-se como função de ligação a função *logit*. Assim, para que a interpretação dos parâmetros ajustados seja facilitada e dada de uma forma mais direta, necessita-se que os parâmetros estimados sejam reescritos na sua escala natural, por meio da função inversa que neste caso é a função *exp*, conforme resultados apresentados na Tabela 16. Neste caso, a função inversa *exp* coincide com as razões de chances estimadas. Os resultados bayesianos são bem próximos dos frequentistas, isso deve-se a utilização de distribuições *a priori* não informativas, porém as estimativas frequentistas estão com intervalos de confiança mais preciso.

Os resultados contidos na Tabela 16, mostram que as sementes de *stevia rebaudiana* quando colocadas na posição 2 tem a mesma chance de ter plântulas normais quando comparado com a posição 1. A razão de chance de 0,77, diz que a chance de produzir plântulas normais sobre a posição 3 é de 0,77 a chance de produzir plântulas normais sobre a posição 1, ou seja, isso é o mesmo que a chance de produzir plântulas normais sobre a posição 1 é  $1/0,77 = 1,30$  vezes maior do que a chance de produzir plântulas normais sobre a posição 3. A luz também está influenciando no tipo de plântulas, pois foi identificado uma razão de chance de 0,79, isso representa que a chance de obter plântulas normais sobre o efeito da luz 2 é equivalente a  $1/0,79 = 1,26$  vezes mais que a chance de obter plântulas normais sobre o efeito da luz 1.

Tabela 16 – Razões de chances (OR) estimadas, intervalos de confiança e e intervalo de credibilidade, referente ao modelo de regressão Binomial e Beta-binomial.

Distribuição	Relação	Estimativas frequentistas		Estimativas bayesianas	
		OR	IC(OR,95%)	OR	IC <sub>r</sub> (OR,95%)
Binomial	Posição 2/Posição 1	0,99	(0,8918; 1,1019)	0,99	(0,8863; 1,0961)
	Posição 3/Posição 1	0,77	(0,6894; 0,8511)	0,77	(0,6851; 0,8467)
	Luz 2/Luz 1	0,79	(0,6894; 0,8511)	0,79	(0,7256; 0,8625)
Beta-binomial	Posição 2/Posição 1	0,99	(0,9055; 1,0770)	0,99	(0,8863; 1,0961)
	Posição 3/Posição 1	0,77	(0,7001; 0,8318)	0,77	(0,6851; 0,8466)
	Luz 2/Luz 1	0,79	(0,7376; 0,8506)	0,79	(0,7256; 0,8625)

No ajuste do modelo de regressão Multinomial também foi utilizado a função de ligação *logit*, e da mesma forma que o modelo de regressão Binomial, a interpretação dos parâmetros ajustados será dada pela função inversa *exp*, e os resultados são apresentados nas Tabelas 17. Neste caso, a função inversa *exp* coincide também com as razões de chances estimadas.

Com relação aos resultados do ajuste do modelo de regressão Multinomial, contido

na Tabela 17, dá-se em duas etapas, primeiro é comparado a proporção de plântulas normais com relação a proporção de sementes não germinadas e em seguida é comparada a proporção de plântulas anormais com relação a proporção de sementes não germinadas. Para a primeira comparação tem-se que a razão de chance de 0,91 compara a chance de produzir plântulas normais na posição 2 com relação a chance de produzir plântulas normais na posição 1, apesar de não ser estatisticamente significativo, tem-se que a chance de produzir plântulas normais na posição 1 é  $1/0,91 = 1,09$  maior que a chance de produzir plântulas normais na posição 2. A chance de uma semente colocada na posição 1 apresentar plântulas normais é equivalente a  $1/0,63 = 1,59$  vezes maior que a chance de uma semente ter plântulas normais na posição 3. Sementes sobre o efeito da luz 1 tem chance de ter plântulas normais que equivale a  $1/0,73 = 1,37$  vezes maior que a chance de obter plântulas normais sobre o efeito da luz 2. Já para a segunda comparação tem-se que a chance de ter plântulas anormais sobre a posição 1 correspondente a  $1/0,76 = 1,32$  vezes mais que a chance de obter sementes anormais na posição 2. A chance de obter sementes anormais sobre a posição 1 é de  $1/0,49 = 2,04$  vezes maior que a chance de obter sementes anormais na posição 3. E por fim, tem-se que sobre o efeito da luz 1 a chance de ter plântulas anormais é de  $1/0,77 = 1,30$  vezes maior que a chance de obter plântulas anormais sobre o efeito da luz 2.

Tabela 17 – Razões de chances estimadas (OR) e os seus respectivos intervalos de confiança do modelo de regressão Multinomial ajustado com relação a proporção de sementes com plântulas normais em relação as sementes não germinadas e a proporção de plântulas anormais com relação as sementes não germinadas.

Logit	Relação	OR	IC(OR,95%)
$\frac{\pi_1}{\pi_3}$	Posição 2/Posição 1	0,91	(0,8173; 1,0101)
	Posição 3/Posição 1	0,63	(0,5682; 0,7015)
$\frac{\pi_2}{\pi_3}$	Luz 2/Luz 1	0,73	(0,6782; 0,8058)
	Posição 2/Posição 1	0,76	(0,6573; 0,8791)
	Posição 3/Posição 1	0,49	(0,4240; 0,5747)
	Luz 2/Luz 1	0,77	(0,6849; 0,8757)

Ao fim, tem-se que ambas as distribuições adotadas são adequadas para se estudar a germinação das sementes de *Stevia rebaudiana*, desta forma a decisão de qual distribuição adotar fica restrita ao interesse do pesquisador. Se o objetivo de interesse for analisar apenas a proporção de plântulas normais em comparação com os outros dois níveis agrupados, deve se utilizar a distribuição Binomial ou Beta-binomial, outro caso em que se pode optar por essas distribuições é quando se observar nos dados que serão analisados uma proporção de plântulas normais ou de sementes não germinadas muito pequena. Por outro lado, a distribuição Multinomial deve ser adotada se o objetivo do pesquisador é entender como se comporta a proporção de plântulas normais, anormais e as sementes não germinadas separadamente. Já com relação a abordagem frequentista e bayesiana verificou-se que a abordagem frequentista retornou intervalos de confiança mais precisos e além disso possibilitou o ajuste com a

distribuição Multinomial, sendo nesse caso a que melhor metodologia para esses dados.



---

## CAPÍTULO 4

---

# CONSIDERAÇÕES FINAIS

---

Os modelos lineares generalizados com efeitos mistos utilizando as distribuições Binomial, Beta-binomial e Multinomial são adequados para explicar o comportamento da germinação das sementes de *Stevia rebaudiana* com relação aos fatores de influência avaliados neste experimento. No entanto, a escolha de um deles para inferir sobre as respostas é uma decisão que pode ser tomada pelo pesquisador de semente. Quando o objetivo for entender apenas a proporção de plântulas normais, o modelo que utilizou a distribuição Binomial e a Beta-binomial são os indicadas, mas se o objetivo for entender todos os três componentes da variável resposta é preciso utilizar o modelo de regressão com distribuição Multinomial. Todos os modelos indicam que a posição da bandeja dentro do germinador e as condições de luz utilizada afetam as chances de obter plântulas normais, anormais ou sementes não germinadas. A posição mais alta dentro do germinador e a luz 2 foram as que mais contribuíram na proporção de plântulas normais de *Stevia rebaudiana*.

---

## REFERÊNCIAS

---

- ABDULLATEEF, R. A.; OSMAN, M. B. Effects of visible light wavelengths on seed germinability in stevia rebaudiana bertonii. *International Journal of Biology*, v. 3, n. 4, p. 83, 2011.
- ABDULLATEEF, R. A.; OSMAN, M. bin; ZAINUDDIN, Z. bint. Acclimatized apparatus enhanced seed germination in stevia rebaudiana bertonii. *International Journal of Biology*, Canadian Center of Science and Education, v. 7, n. 2, p. 28, 2015.
- AGRICULTURA, P. e. A. Ministério da. *Regras para análise de sementes*. [S.l.]: MAPA/ACS Brasília, 2009.
- ARORA, E. et al. Stevia: A promising herbal sweeteners. *JK SCIENCE Oct-December*, v. 12, n. 4, 2010.
- BOLSTAD, W. M.; CURRAN, J. M. *Introduction to Bayesian statistics*. [S.l.]: John Wiley & Sons, 2016.
- BOX, G. E.; TIAO, G. C. *Bayesian inference in statistical analysis*. [S.l.]: John Wiley & Sons, 2011. v. 40.
- CARNEIRO, J.; GUEDES, T. Influência do contato das sementes de stevia (stevia rebaudiana (bert.) bertonii), no substrato, avaliada pela função de weibull. *Revista Brasileira de Sementes*, v. 4, n. 1, p. 65–68, 1992.
- CARNEIRO, J. W. P. Determinação do número de sementes para avaliar o desempenho germinativo de stevia (stevia rebaudiana (bert.) bertonii). *Revista Brasileira de Sementes*, v. 18, n. 01, p. 1–5, 1996.
- CARNEIRO, J. W. P. Stevia rebaudiana (bert.) bertonii: Stages of plant development. *Canadian Journal of Plant Science*, NRC Research Press, v. 87, n. 4, p. 861–865, 2007.
- CASELLA, G.; BERGER, R. L. *Statistical inference*. [S.l.]: Duxbury Pacific Grove, CA, 2002. v. 2.
- CHRISTENSEN, R. et al. *Bayesian ideas and data analysis: an introduction for scientists and statisticians*. [S.l.]: CRC Press, 2011.
- COLE, T. J.; GREEN, P. J. Smoothing reference centile curves: the lms method and penalized likelihood. *Statistics in medicine*, Wiley Online Library, v. 11, n. 10, p. 1305–1319, 1992.

- DEGROOT, M. H. M. H. et al. *Probability and statistics*. [S.l.: s.n.], 1986.
- DONOGHOE, M. W.; MARSCHNER, I. C. Flexible regression models for rate differences, risk differences and relative risks. *The international journal of biostatistics*, v. 11, n. 1, p. 91–108, 2015.
- DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, Taylor & Francis, v. 5, n. 3, p. 236–244, 1996.
- GELFAND, A. E.; DEY, D. K. Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 501–514, 1994.
- GNEITING, T.; RAFTERY, A. E. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, Taylor & Francis, v. 102, n. 477, p. 359–378, 2007.
- ISTA. *International rules for seed testing*. [S.l.]: International Seed Testing Association, Bassersdorf, Switzerland, 2013.
- LEHMANN, E. L.; CASELLA, G. *Theory of point estimation*. [S.l.]: Springer Science & Business Media, 2006.
- LEMUS-MONDACA, R. et al. Stevia rebaudiana bertonii, source of a high-potency natural sweetener: A comprehensive review on the biochemical, nutritional and functional aspects. *Food Chemistry*, Elsevier, v. 132, n. 3, p. 1121–1132, 2012.
- MAGALHÃES, M. N. *Probabilidade e variáveis aleatórias*. [S.l.]: Edusp, 2006.
- MORETTIN, L. G. *Estatística básica: probabilidade e inferência: volume único*. [S.l.]: Pearson Prentice Hall, 2010.
- NASSIF, S. M. L.; VIEIRA, I. G.; FERNADES, G. D. *Fatores Externos (ambientais) que Influenciam na Germinação de Sementes*. 1998. Disponível em: <<http://www.ipef.br/tecsementes/germinacao.asp>>. Acesso em: 10 mar. 2016.
- NELDER, J. A.; BAKER, R. J. Generalized linear models. *Encyclopedia of statistical sciences*, Wiley Online Library, 1972.
- ÖZYIĞİT, Y.; ESRA, U.; TURGUT, K. The effect of different pollination methods on seed yield and germination features in stevia rebaudiana bertonii. *Türkiye Tarımsal Araştırmalar Dergisi*, v. 2, n. 2, p. 114–117, 2015.
- PAULA, G. A. *Modelos de regressão: com apoio computacional*. [S.l.]: IME-USP São Paulo, 2004.
- PAWITAN, Y. *In all likelihood: statistical modelling and inference using likelihood*. [S.l.]: Oxford University Press, 2001.
- POTVIN, C. Anova: experiments in controlled environments. *Design and analysis of ecological experiments*. Chapman and Hall, New York, New York, USA, p. 46–68, 1993.
- R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2016. Disponível em: <<https://www.R-project.org/>>.

- RESENDE, M. D. V. de. *Análise estatística de modelos mistos via REML/BLUP na experimentação em melhoramento de plantas perenes*. [S.l.]: Embrapa Florestas, 2000.
- RIGBY, R.; STASINOPOULOS, D. A semi-parametric additive model for variance heterogeneity. *Statistics and Computing*, Springer, v. 6, n. 1, p. 57–65, 1996.
- RIGBY, R.; STASINOPOULOS, D. *A flexible regression approach using GAMLSS in R*, 2010. 2012.
- RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Wiley Online Library, v. 54, n. 3, p. 507–554, 2005.
- ROSENBLATT, M. Remarks on a multivariate transformation. *The annals of mathematical statistics*, JSTOR, v. 23, n. 3, p. 470–472, 1952.
- RUE, H.; MARTINO, S.; CHOPIN, N. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, Wiley Online Library, v. 71, n. 2, p. 319–392, 2009.
- SIMLAT, M. et al. The effect of light quality on seed germination, seedling growth and selected biochemical properties of stevia rebaudiana bertonii. *Scientia Horticulturae*, Elsevier, v. 211, p. 295–304, 2016.
- SORENSEN, D.; GIANOLA, D. *Likelihood, Bayesian, and MCMC methods in quantitative genetics*. [S.l.]: Springer Science & Business Media, 2007.
- SPIEGELHALTER, D. J. et al. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 64, n. 4, p. 583–639, 2002.
- UÇAR, E.; ÖZYİĞİT, Y.; TURGUT, K. The effects of light and temperature on germination of stevia (stevia rebaudiana bert.) seeds. 2016.
- VEHTARI, A.; GELMAN, A. Waic and cross-validation in stan. *Submitted*. [http://www.stat.columbia.edu/~gelman/research/unpublished/waic\\_stan.pdf](http://www.stat.columbia.edu/~gelman/research/unpublished/waic_stan.pdf) Accessed, v. 27, n. 2015, p. 5, 2014.
- WATANABE, S. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, v. 11, n. Dec, p. 3571–3594, 2010.

---

## ANEXO A

---

# CÓDIGOS INFERÊNCIA FREQUENTISTA

---

### Modelo Binomial

Modelo 0:  $y = \beta_0$

```
mod0 <- gamlss(formula = y ~ 1,family = BI(mu.link = "logit"), data = dados1)
summary(mod0)
plot(mod0); wp(mod0)
AIC = GAIC(mod0, k= 2)
AICc = GAIC(mod0, c = T)
DG = GAIC(mod0, k = 0)
BIC = GAIC(mod0, k = log(nrow(dados1)))
```

Modelo 1:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3$

```
mod1 <- gamlss(formula = y ~ Pos2+Pos3,family = BI(mu.link = "logit"), data = dados1)
summary(mod1)
plot(mod1); wp(mod1)
AIC = GAIC(mod1, k= 2)
AICc = GAIC(mod1, c = T)
DG = GAIC(mod1, k = 0)
BIC = GAIC(mod1, k = log(nrow(dados1)))
```

Modelo 2:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2$

```
mod2<- gamlss(formula = y ~ Posicao + Luz , family = BI(mu.link = "logit"), data =
dados1)
```

```
summary(mod2)
plot(mod2); wp(mod2)
AIC = GAIC(mod2, k= 2)
AICc = GAIC(mod2, c = T)
DG = GAIC(mod2, k = 0)
BIC = GAIC(mod2, k = log(nrow(dados1)))
```

Modelo 3:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Lote$

```
mod3 <- gamlss(formula = y ~ Posicao+Luz+ re(random = 1|Lote,method = "REML"),
family = BI(mu.link = "logit"), data = dados1)
```

```
res = summary(mod3)
plot(mod3); wp(mod3)
AIC = GAIC(mod3, k= 2)
AICc = GAIC(mod3, c = T)
DG = GAIC(mod3, k = 0)
BIC = GAIC(mod3, k = log(nrow(dados1)))
```

### Método delta e o cálculo do intervalo de confiança da razão de chance.

```
MDe = exp(res[,1])*res[,2]
LI = exp(res[,1])-qnorm(p = .975)*MDe
LS = exp(res[,1])+qnorm(p = .975)*MDe
```

Modelo 4:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Germinador$

```
mod4 <- gamlss(formula = y ~ Posicao + Luz + re(random = ~ 1|Germinador, method
= "REML"), family = BI(mu.link = "logit"), data = dados1)
```

```
summary(mod4)
plot(mod4); wp(mod4)
AIC = GAIC(mod4, k= 2)
AICc = GAIC(mod4, c = T)
DG = GAIC(mod4, k = 0)
BIC = GAIC(mod4, k = log(nrow(dados1)))
```

Modelo 5:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Germinador + b_2 Lote$

```
mod5 <- gamlss(formula = y ~ Posicao+Luz+ re(random = ~ 1|Germinador,method =
"REML")+ re(random = ~ 1|Lote,method = "REML"),family = BI(mu.link = "logit"), data =
dados1)
```

```
summary(mod5)
```

```
plot(mod5); wp(mod5)
```

```
AIC = GAIC(mod5, k= 2)
```

```
AICc = GAIC(mod5, c = T)
```

```
DG = GAIC(mod5, k = 0)
```

```
BIC = GAIC(mod5, k = log(nrow(dados1)))
```

## Modelo Beta-binomial

Modelo 0:  $y = \beta_0$

```
mod0 <- gamlss(formula = y ~ 1, family = BB(mu.link = "logit",sigma.link = "log"), data
= dados1) summary(mod0)
```

```
plot(mod0); wp(mod0)
```

```
AIC = GAIC(mod0, k= 2)
```

```
AICc = GAIC(mod0, c = T)
```

```
DG = GAIC(mod0, k = 0)
```

```
BIC = GAIC(mod0, k = log(nrow(dados1)))
```

Modelo 1:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3$

```
mod1 <- gamlss(formula = y ~ Posicao, family = BB(mu.link = "logit",sigma.link = "log"),
data = dados1)
```

```
summary(mod1)
```

```
plot(mod1); wp(mod1)
```

```
AIC = GAIC(mod1, k= 2)
```

```
AICc = GAIC(mod1, c = T)
```

```
DG = GAIC(mod1, k = 0)
```

```
BIC = GAIC(mod1, k = log(nrow(dados1)))
```

Modelo 2:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2$

```
mod1 <- gamlss(formula = y ~ Posicao + Luz, family = BB(mu.link = "logit",sigma.link
= "log"), data = dados1) summary(mod2)
plot(mod2); wp(mod2)
AIC = GAIC(mod2, k= 2)
AICc = GAIC(mod2, c = T)
DG = GAIC(mod2, k = 0)
BIC = GAIC(mod2, k = log(nrow(dados1)))
```

Modelo 3:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Lote$

```
mod3 <- gamlss(formula = y ~ Posicao + Luz + re(random = ~ 1|Lote, method =
"REML"), family = BB(mu.link = "logit", sigma.link = "log"), data = dados1)
res = summary(mod3)
plot(mod3); wp(mod3)
AIC = GAIC(mod3, k= 2)
AICc = GAIC(mod3, c = T)
DG = GAIC(mod3, k = 0)
BIC = GAIC(mod3, k = log(nrow(dados1)))
```

### Método delta e o cálculo do intervalo de confiança da razão de chance.

Modelo 4:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Germinador$

```
mod4 <- gamlss(formula = y ~ Posicao + Luz+ re(random = ~ 1|Germinador, method =
"REML"), family = BB(mu.link = "logit", sigma.link = "log"), data = dados1) summary(mod4)
plot(mod4); wp(mod4)
AIC = GAIC(mod4, k= 2)
AICc = GAIC(mod4, c = T)
DG = GAIC(mod4, k = 0)
BIC = GAIC(mod4, k = log(nrow(dados1)))
```

Modelo 5:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Germinador + b_2 Lote$

```
mod5 <- gamlss(formula = y ~ Posicao + Luz + re(random = ~ 1|Germinador, method =
"REML") + re(random = ~ 1|Lote, method = "REML"), family = BB(mu.link = "logit", sigma.link
= "log"), data = dados1) summary(mod5)
```



```
plot(mod5); wp(mod5)
AIC = GAIC(mod5, k= 2)
AICc = GAIC(mod5, c = T)
DG = GAIC(mod5, k = 0)
BIC = GAIC(mod5, k = log(nrow(dados1)))
```

## Modelo Multinomial

Modelo 0:  $y = \beta_0$

```
mod0 <- gamlss(formula = Y ~ 1, sigma.formula = Y ~ 1, family = MULTIN(type = "3"),
data = dados1)
summary(mod0)
plot(mod0); wp(mod0)
AIC = GAIC(mod0, k= 2)
AICc = GAIC(mod0, c = T)
DG = GAIC(mod0, k = 0)
BIC = GAIC(mod0, k = log(nrow(dados1)))
```

Modelo 1:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3$

```
mod1 <- gamlss(formula = Y ~ Posicao, sigma.formula = Y ~ Posicao, family = MUL-
TIN(type = "3"), data = dados1)
summary(mod1)
plot(mod1); wp(mod1)
AIC = GAIC(mod1, k= 2)
AICc = GAIC(mod1, c = T)
DG = GAIC(mod1, k = 0)
BIC = GAIC(mod1, k = log(nrow(dados1)))
```

Modelo 2:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2$

```
mod2 <- gamlss(formula = Y ~ Posicao+Luz, sigma.formula = Y ~ Posicao+Luz, family
= MULTIN(type = "3"), data = dados1)
summary(mod2)
plot(mod2); wp(mod2)
```

```

AIC = GAIC(mod2, k= 2)
AICc = GAIC(mod2, c = T)
DG = GAIC(mod2, k = 0)
BIC = GAIC(mod2, k = log(nrow(dados1)))

```

Modelo 3:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Lote$

```

mod3 <- gamlss(formula = Y ~ Posicao+Luz+re(random = ~1 | Lote), sigma.formula =
Y ~ Posicao+Luz+re(random = ~1 | Lote), family = MULTIN(type = "3"), data = dados1)

```

```

res = summary(mod3)
plot(mod3); wp(mod3)
AIC = GAIC(mod3, k= 2)
AICc = GAIC(mod3, c = T)
DG = GAIC(mod3, k = 0)
BIC = GAIC(mod3, k = log(nrow(dados1)))

```

### Método delta e o cálculo do intervalo de confiança da razão de chance.

```

IC=data.frame(LI=res[,1]-qnorm(p = .975)*res[,2],LS=res[,1]+qnorm(p = .975)*res[,2])

```

Modelo 4:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Germinador$

```

mod4 <- gamlss(formula = Y ~ Posicao+Luz+re(random = ~ 1 | Germinador), sigma.formula
=Y ~ Posicao+Luz+re(random = ~ 1 | Germinador), family = MULTIN(type = "3"), data=dados1)

```

```

summary(mod4)
plot(mod4); wp(mod4)
AIC = GAIC(mod4, k= 2)
AICc = GAIC(mod4, c = T)
DG = GAIC(mod4, k = 0)
BIC = GAIC(mod4, k = log(nrow(dados1)))

```

Modelo 5:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Germinador + b_2 Lote$

```

mod4 <- gamlss(formula = Y ~ Posicao+Luz+re(random = ~ 1 | Germinador)+re(random
= 1|Lote), sigma.formula =Y ~ Posicao+Luz+re(random = ~ 1 | Germinador)+re(random =
1|Lote), family = MULTIN(type = "3"), data=dados1)

```

```

summary(mod5)
plot(mod5); wp(mod5)

```

AIC = GAIC(mod5, k= 2)

AICc = GAIC(mod5, c = T)

DG = GAIC(mod5, k = 0)

BIC = GAIC(mod5, k = log(nrow(dados1)))

---

## ANEXO B

---

# CÓDIGOS INFERÊNCIA BAYESIANA

---

### Binomial

Modelo 0:  $y = \beta_0$

```
fit0 = inla(y 1, data=dados1, Ntrials=n, family="binomial",control.compute = list(dic = TRUE, waic = TRUE,cpo=TRUE))
```

```
summary(fit0)
```

```
plot(fit0)
```

```
DIC = fit0$dic$dic; WAIC = fit0$waic$waic; LPML= sum(log(fit0$cpo$cpo))
```

Modelo 1:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3$

```
fit1 = inla(y factor(Posicao), data=dados1, Ntrials=n, family="binomial",control.compute = list(dic = TRUE, waic = TRUE,cpo=TRUE))
```

```
summary(fit1)
```

```
plot(fit1)
```

```
DIC = fit1$dic$dic; WAIC = fit1$waic$waic; LPML= sum(log(fit1$cpo$cpo))
```

Modelo 2:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2$

```
fit2 = inla(y factor(Posicao)+factor(Luz),data=dados1, Ntrials=n, family="binomial",control.compute = list(dic = TRUE, waic = TRUE,cpo=T))
```

```
summary(fit2)
```

```
plot(fit2)
```

```
DIC = fit2$dic$dic; WAIC = fit2$waic$waic; LPML= sum(log(fit2$cpo$cpo))
```

Modelo 3:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Lote$

```
fit3 = inla(y factor(Posicao)+factor(Luz)+f(Lote,model="iid",param=c(0.001,0.001)) , data=dados1,
Ntrials=n, family="binomial",control.compute = list(dic = TRUE, waic = TRUE,cpo=T))
summary(fit3)
plot(fit3)
DIC = fit3$dic$dic; WAIC = fit3$waic$waic; LPML= sum(log(fit3$cpo$cpo))
```

Modelo 4:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Germinador$

```
fit4 = inla(y factor(Posicao)+factor(Luz)+f(Germinador,model="iid",param=c(0.001,0.001))
, data=dados1, Ntrials=n, family="binomial",control.compute = list(dic = TRUE, waic = TRUE,cpo=T))
summary(fit4)
plot(fit4)
DIC = fit4$dic$dic; WAIC = fit4$waic$waic; LPML= sum(log(fit4$cpo$cpo))
```

Modelo 5:  $y = \beta_0 + \beta_1 Pos2 + \beta_2 Pos3 + \beta_3 Luz2 + b_1 Germinador + b_2 Lote$

```
fit5 = inla(y factor(Posicao)+factor(Luz)+f(Lote,model="iid",param=c(0.001,0.001)) +f(Germinador,m
data=dados1, Ntrials=n,
family="binomial",control.compute = list(dic = TRUE, waic = TRUE,cpo=T))
summary(fit5)
plot(fit5)
DIC = fit5$dic$dic; WAIC = fit5$waic$waic; LPML= sum(log(fit5$cpo$cpo))
```

## Beta-binomial

O ajuste dos modelos utilizando a distribuição Beta-binomial se da como os mesmo comandos utilizados na Binomial, a única coisa que muda é a troca de family="binomial"por family="betabinomial".