**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Classificação de variedades de cana-de-açúcar quanto à suscetibilidade
a carvão e mosaico para dados com superdispersão e excesso de zeros**

Sandra Elisa Fonrozo

Dissertação apresentada para obtenção do título de
Mestra em Ciências. Área de concentração: Estatística e Experimentação Agronômica

**Piracicaba
2021**

Sandra Elisa Fonrozo
Bacharelado em Estatística

**Classificação de variedades de cana-de-açúcar quanto à suscetibilidade
a carvão e mosaico para dados com superdispersão e excesso de zeros**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Profª Drª **CLARICE GARCIA BORGES DEMÉ-
TRIO**

Dissertação apresentada para obtenção do título de
Mestra em Ciências. Área de concentração: Estatís-
tica e Experimentação Agronômica

Piracicaba
2021

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Fonrozo, Sandra Elisa

Classificação de variedades de cana-de-açúcar quanto à suscetibilidade a carvão e mosaico para dados com superdispersão e excesso de zeros / Sandra Elisa Fonrozo. -- versão revisada de acordo com a resolução CoPGr 6018 de 2011. -- Piracicaba, 2021 .

89 p.

Dissertação (Mestrado) -- USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Incidência 2. Proporção 3. Superdispersão 4. Fitopatologia 5. Triagem de doenças 6. Classificação de variedades. . I. Título.

DEDICATÓRIA

*A **Deus**, que me abençoa a cada dia, me proporcionando muitas experiências de vida, aprendizados e colocando em meu caminho vários anjos que me ajudam a evoluir e ser feliz.*

*Aos meus pais **Oswaldo** e **Selma**, que sempre me ensinaram a batalhar e trabalhar de forma digna e com respeito a todos. Sempre deram o melhor de si e, com muito carinho e dedicação, nunca mediram esforços para que eu chegasse até aqui. Tudo o que sou hoje é resultado da educação, amor e apoio deles, pois sem eles este trabalho e muitos outros sonhos não se realizariam.*

*Ao meu amado esposo **Felipe** que está sempre ao meu lado proporcionando muito apoio, amor, companheirismo e compreensão.*

*Às minhas irmãs **Solange** e **Simone**, que sempre se preocupam com meu bem estar, com muito carinho e incentivo.*

*Às minhas sobrinhas **Sofia** e **Sarah**, que com tanta leveza e doçura, trazem muita alegria para minha vida.*

A todos os professores que passaram em minha vida, deixando sempre tanta contribuição.

E a todos os meus familiares, amigos e colegas de trabalho.

AGRADECIMENTOS

A Deus, pela dádiva da vida e por me permitir realizar tantos sonhos, inclusive, a finalização desse trabalho.

À minha mãe e ao meu pai deixo um agradecimento especial, por todas as lições de amor, companheirismo, amizade, caridade, dedicação, respeito e compreensão que vocês me dão a cada dia. Sou privilegiada por ter pais tão especiais. Às minhas irmãs queridas, sempre prontas a me apoiar em tudo nesta vida.

Ao meu amado esposo, por todo amor, carinho, compreensão e apoio em tantos momentos difíceis desta caminhada. Obrigado por permanecer ao meu lado, em todos os momentos da minha vida.

À Prof^{ra} Clarice Garcia Borges Demétrio, pela orientação, competência e profissionalismo. Obrigada por acreditar em mim e por toda preocupação despendida.

Aos membros da banca examinadora, Prof^{ra} Liciania Vaz de Arruda Silveira, Prof. Cristian Marcelo Villegas Lobos e Prof^{ra} Mariana Ragassi Urbano, que aceitaram participar e colaborar com esta dissertação. À Prof^{ra} Giovana Fumes Ghantous, agradeço ainda por toda ajuda e conversas, que foram importantíssimas para a finalização desse trabalho. Ao professor Rafael De Andrade Moral por toda ajuda no entendimento das funções do R.

A todos os meus professores do ICMC/USP que tanto me ensinaram. Em especial ao Prof. Mário de Castro Andrade Filho, exemplo de professor, muito dedicado a seus alunos.

A todos os meus professores da ESALQ/USP que proporcionaram ainda mais conhecimento sobre estatística e experimentação.

A todos os meus colegas de mestrado, que encararam comigo tantos desafios.

A Solange Sabadin, mais que secretária, esteve sempre pronta para me ajudar.

Ao Centro de Tecnologia Canavieira (CTC) por tantas oportunidades de crescimento profissional e pessoal. Em especial ao Michael Keith Butterfield, que sempre me surpreende com tanto conhecimento e experiência de vida. Obrigada por todos os ensinamentos e oportunidades.

A todos os meus colegas do CTC, e em especial, a Aline Cristina Zavaglia e toda equipe de Fitopatologia que me ensinaram tanto sobre doença de cana-de-açúcar, com muita paciência e dedicação.

A todos, que contribuíram de forma direta ou indireta para a realização deste trabalho, o meu muito obrigada.

EPÍGRAFE

“A mente que se abre a uma nova ideia jamais voltará ao seu tamanho original”.

Albert Einstein

SUMÁRIO

Resumo	8
Abstract	9
Lista de Figuras	10
Lista de Tabelas	13
1 Introdução	15
1.1 Carvão	15
1.2 Mosaico	16
1.3 Justificativa e importância do trabalho	16
1.4 Nomenclatura	18
2 Revisão bibliográfica	19
2.1 Modelos lineares generalizados (MLG)	19
2.1.1 Estimação dos parâmetros	20
2.1.2 Função desvio e estatística X^2 de Pearson	20
2.1.3 Análise do desvio e seleção de modelos	21
2.1.4 Estimação do parâmetro de dispersão	22
2.1.5 Técnicas para diagnósticos em MLG	23
2.1.5.1 Tipos de resíduos	23
2.1.5.2 Técnicas gráficas	24
2.1.5.3 Verificação da função de ligação	24
2.2 Modelos para dados de proporção	25
2.2.1 Modelo binomial	25
2.2.2 Superdispersão	25
2.2.3 Modelo quase-binomial	27
2.2.4 Modelo beta-binomial	28
2.2.5 Modelos inflacionados de zeros	28
2.2.5.1 Modelo binomial inflacionado de zeros - ZIBI	28
2.2.5.2 Modelo beta-binomial inflacionado de zeros - ZIBB	29
2.3 Pacote GAMLSS	29
3 Material e Métodos	31
3.1 Material	31
3.1.1 Preparação das mudas e inoculação de carvão	31
3.1.2 Linha infectora do mosaico	32
3.1.3 Experimentos de campo	32
3.2 Métodos	33
3.2.1 Análises exploratórias	33
3.2.2 Modelagem	33
3.2.3 Agrupamento de variedades	35
3.2.4 Classificação dos grupos de variedades	36
4 Resultados e Discussões	39
4.1 Estrutura dos dados	39
4.2 Análise exploratória	41
4.3 Modelagem	46
4.4 Agrupamento de variedades	54
4.5 Classificação dos grupos de variedades	57

5	Considerações finais	69
	Referências	71
	Apêndices	73

RESUMO

Classificação de variedades de cana-de-açúcar quanto à suscetibilidade a carvão e mosaico para dados com superdispersão e excesso de zeros

Grande parte do sucesso da cana-de-açúcar no mundo deve-se aos programas de melhoramento genético que selecionam as melhores variedades baseados, principalmente, em características de produtividade e resistência a doenças. O carvão e o mosaico são doenças que causam grandes prejuízos em canaviais e o principal controle se dá pelo uso de variedades resistentes. Para carvão e mosaico, a observação da incidência da doença é feita pela proporção de plantas sintomáticas e, devido a sua natureza, esses dados não seguem uma distribuição normal, sendo essa uma pressuposição estabelecida nos modelos lineares clássicos.

Dessa forma, o objetivo desse trabalho é apresentar uma metodologia de classificação de variedades de cana-de-açúcar quanto à suscetibilidade, para as doenças carvão e mosaico, utilizando modelos lineares generalizados e extensões ajustados para proporção de plantas sintomáticas. A metodologia foi aplicada a dois ensaios avaliados em dois estágios de corte e, pela natureza dos dados, houve problemas de superdispersão e excesso de zeros, resultando na necessidade da utilização de modelos mais elaborados. Dentro da classe de modelos lineares generalizados e extensões, foram abordados os modelos binomial; quasi-binomial e beta-binomial (adequação de dados com superdispersão); binomial inflacionado de zeros e beta-binomial inflacionados de zeros (adequação de dados com excesso de zeros). Os agrupamentos de variedades foram realizados a partir do teste de razão de verossimilhanças para modelos encaixados e critério de informação de Akaike (AIC) e, a classificação de variedades, baseada nos resultados das variedades de referência (padrões). Os resultados mostraram ser satisfatórios para classificação das variedades quanto à suscetibilidade a carvão e mosaico, desde que as variedades de referência dos ensaios sejam representativas.

Todas as análises foram realizadas no software R. Para ajuste dos modelos utilizou-se o pacote *glm* para os modelos binomial e quase-binomial, e os demais modelos foram ajustados pelo pacote *gamlss*.

Palavras-chave: Incidência, Proporção, Superdispersão, Fitopatologia, Triagem de doenças, Classificação de variedades.

ABSTRACT

Classification of sugarcane varieties according to susceptibility to smut and mosaic for data with overdispersion and excess of zeros

Much of the success of sugarcane in the world is due to Genetic Improvement Programs that select the best varieties based in characteristics of, mainly, productivity and resistance to diseases. Smut and Mosaic are diseases that cause great damage in the canebrake and main control is done by resistant varieties. To smut and mosaic, the incidence of disease is evaluated for proportion of symptomatic plants and, because of their nature, these data do not follow a normal distribution, which is a pre assumption established in classical linear models.

Thus, the objective of this work is to present a methodology for classifying sugarcane varieties in terms of susceptibility to smut and mosaic diseases, using generalized linear models and extensions fitted to the proportion of symptomatic plants. The methodology was applied to two trials evaluated for two harvests and, because of the nature of the data, there were problems of overdispersion and excess of zeros, being necessary to use more elaborated models. Within the class of generalized linear models and extensions, were addressed binomial models; quasi-binomial and beta-binomial (adequacy for overdispersion data); zero-inflated binomial and zero-inflated beta-binomial (adequacy for data with excess of zeros). Variety groupings were performed using the likelihood ratio to nested models and Akaike's information criterion (AIC) and the variety classification was based on the results of reference varieties (standards). The results proved to be satisfactory for classifying the varieties for smut and mosaic susceptibility, as long as the reference varieties of the trials are representative.

All analyzes were performed using the R software. To fit the models, the *glm* package was used for the binomial and quasi-binomial models, and the other models were fitted by the *gamlss* package.

Keywords: Incidence, Proportion, Overdispersion, Plant pathology, Disease Screening Trials, Variety classification.

LISTA DE FIGURAS

1.1	Triângulo da doença (CARDOSO, 2017)	15
1.2	Chicote de carvão em cana-de-açúcar	16
1.3	Mosaico da cana-de-açúcar	16
1.4	Ciclo da cana	18
1.5	Representação de toletes de cana-de-açúcar à esquerda e gema à direita	18
2.1	Representação da função de ligação <i>logit</i>	25
2.2	Distribuição da variância da distribuição de probabilidade binomial	26
3.1	1) Distribuição dos toletes nas bandejas 2) Inoculação do carvão 3) Bandejas inoculadas, cobertas com substrato 4) Plantas após 30 dias.	31
3.2	Esquema da linha infectora de mosaico nos ensaios de campo	32
4.1	Distribuição da incidência de carvão dos ensaios e cortes, sem variedades com variância nula.	40
4.2	Distribuição da incidência de mosaico dos ensaios e cortes, sem variedades com variância nula.	41
4.3	Distribuição da incidência média de mosaico e carvão para ensaios 2016 e 2017, por categoria e estágio de corte	42
4.4	Distribuição da incidência de carvão (ensaio de 2016). As variedades foram ordenadas pela média de incidência de carvão e as cores representam: clones(cinza), padrões resistentes(verde), padrões intermediários(amarelo) e padrões suscetíveis(vermelho). As variedades com variância nula em ambos os cortes foram desconsideradas.	43
4.5	Distribuição da incidência média de carvão (ensaio de 2017). As variedades foram ordenadas pela média de incidência de carvão e as cores representam: clones(cinza), padrões resistentes(verde), padrões intermediários(amarelo) e padrões suscetíveis(vermelho). As variedades com variância nula em ambos os cortes foram desconsideradas.	43
4.6	Distribuição da incidência média de mosaico (ensaio de 2016). As variedades foram ordenadas pela média de incidência de mosaico e as cores representam: clones(cinza), padrões resistentes(verde), padrões intermediários(amarelo) e padrões suscetíveis(vermelho). As variedades com variância nula em ambos os cortes foram desconsideradas.	43
4.7	Distribuição da incidência média de mosaico (ensaio de 2017). As variedades foram ordenadas pela média de incidência de mosaico e as cores representam: clones(cinza), padrões resistentes(verde), padrões intermediários(amarelo) e padrões suscetíveis(vermelho). As variedades com variância nula em ambos os cortes foram desconsideradas.	44
4.8	Representação da função de ligação <i>logit</i> à esquerda e função de variância, à direita	44
4.9	Incidência média de carvão e interação dos dois cortes dos padrões de referência.	44
4.10	Incidência média de mosaico e interação dos dois cortes dos padrões de referência.	45
4.11	Dispersão entre a incidência média de carvão e mosaico em cana planta e soca (ensaios 2016 e 2017). A linha diagonal representa a função identidade.	45
4.12	Gráficos semi-normais para modelos binomiais ajustados para incidência de carvão à esquerda, e mosaico à direita. Ajuste agrupando variedades da Tabela 4.7	47
4.13	Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo binomial inflacionado (ZIBI) ajustado para incidência de carvão (ensaio de 2016 - cana planta)	50
4.14	Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo beta-binomial (BB) ajustado para incidência de mosaico (ensaio de 2016 - cana planta)	50

4.15	Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo beta-binomial inflacionado (ZIBB) ajustado para incidência de carvão (ensaio de 2016 - cana soca)	51
4.16	Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo beta-binomial (BB) ajustado para incidência de mosaico (ensaio de 2016 - cana soca)	51
4.17	Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo binomial (BI) ajustado para incidência de carvão (ensaio de 2017 - cana planta)	52
4.18	Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo binomial (BI) ajustado para incidência de mosaico (ensaio de 2017 - cana planta)	52
4.19	Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo beta-binomial (BB) ajustado para incidência de carvão (ensaio de 2017 - cana soca)	53
4.20	Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo beta-binomial (BB) ajustado para incidência de mosaico (ensaio de 2017 - cana soca)	53
4.21	Comparação das funções de variância estimadas pelos modelos finais	54
4.22	Representação do esquema de agrupamento sequencial das variedades ordenadas pela incidência predita em forma decrescente, por teste de razão de verossimilhança para modelos encaixados.	56
4.23	Representação dos ajustes nos grupos formados, para escolha daquele que tiver menor <i>deviance</i> residual.	56
4.24	Representação dos grupos de variedades gerados por modelos ZIBI encaixados para dados de carvão para o ensaio de 2016 (cana planta). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.	59
4.25	Representação dos grupos de variedades gerados por modelos ZIBB encaixados para dados de carvão para o ensaio de 2016 (cana soca). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.	60
4.26	Representação dos grupos de variedades gerados por modelos BI encaixados para dados de carvão para o ensaio de 2017 (cana planta). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.	61
4.27	Representação dos grupos de variedades gerados por modelos BB encaixados para dados de carvão para o ensaio de 2017 (cana soca). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.	62
4.28	Representação dos grupos de variedades gerados por modelos BB encaixados para dados de mosaico para o ensaio de 2016 (cana planta). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.	63

4.29	Representação dos grupos de variedades gerados por modelos BB encaixados para dados de mosaico para o ensaio de 2016 (cana soca). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.	64
4.30	Representação dos grupos de variedades gerados por modelos BI encaixados para dados de mosaico para o ensaio de 2017 (cana planta). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.	65
4.31	Representação dos grupos de variedades gerados por modelos BB encaixados para dados de mosaico para o ensaio de 2017 (cana soca). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.	66
4.32	% de variedades com classificação equivalente em cana planta e cana soca, para dados de carvão e mosaico dos ensaios de 2016 e 2017.	66
A.1	Gráficos semi-normais para modelos Binomiais logísticos ajustados para carvão	77
A.2	Gráficos semi-normais para modelos Quase-binomiais ajustados para carvão	78
A.3	Gráficos semi-normais para modelos Beta-binomiais ajustados para carvão	78
A.4	Gráficos semi-normais para modelos Binomiais inflacionados ajustados para carvão	79
A.5	Gráficos semi-normais para modelos Beta-binomiais inflacionados ajustados para carvão	79
A.6	Gráficos semi-normais para modelos Binomiais logísticos ajustados para mosaico	80
A.7	Gráficos semi-normais para modelos Quase-binomiais ajustados para mosaico	80
A.8	Gráficos semi-normais para modelos Beta-binomiais ajustados para mosaico	81
A.9	Gráficos semi-normais para modelos Binomiais inflacionados ajustados para mosaico	81
A.10	Gráficos semi-normais para modelos Beta-binomiais inflacionados ajustados para mosaico	82

LISTA DE TABELAS

1.1	Caracterização de variedades quanto à resistência a doenças, seguindo a escala de HUTCHINSON e DANIELS (1971)	17
1.2	Classificação de variedades seguindo a escala de CHAO et al. (1989)	17
2.1	Parametrização dos modelos no pacote GAMLSS	29
3.1	Descrição dos ensaios	32
3.2	Classificação histórica dos padrões	33
3.3	Descrição das funções de ligação utilizadas nos modelos, seguindo a parametrização do pacote <i>gamlss</i>	35
3.4	Classificação dos grupos de variedades gerados pelos modelos encaixados, dada pela menor classe histórica dos padrões presentes no grupo	37
3.5	Classificação dos grupos de variedades, dada pela menor classe histórica dos padrões presentes no grupo, com aplicação dos limites estabelecidos pela escala de CHAO et al. (1989).	37
4.1	Situação real das parcelas dos ensaios	39
4.2	Porcentagem de zeros observados nos conjuntos de dados	39
4.3	Número de variedades consideradas nos modelos	40
4.4	Porcentagem de zeros observados nos conjuntos de dados, sem as variedades com variância nula	40
4.5	Estatísticas do modelo Binomial logístico para incidência de carvão	46
4.6	Estatísticas do modelo Binomial logístico para incidência de mosaico	46
4.7	Lista de variedades com apenas uma planta sintomática em apenas uma das repetições	47
4.8	Estatísticas dos modelos ajustados para incidência de carvão	48
4.9	Estatísticas dos modelos ajustados para incidência de mosaico	49
4.10	Estimativas dos parâmetros do modelo BB completo para ensaio de 2017 - soca (mosaico)	55
4.11	Estimativas dos parâmetros do modelo BB para ensaio de 2017 - soca (mosaico)	57
4.12	Estatísticas dos modelos completo (por variedade) e reduzido (por grupo de variedades)	57
4.13	Atribuição das classes de resistência a carvão para grupos do ensaio de 2016 (cana planta), em que R: resistente, I: intermediário e S: suscetível	58
4.14	Atribuição das classes de resistência a carvão para grupos do ensaio de 2016 (cana soca), em que R: resistente, I: intermediário e S: suscetível	59
4.15	Atribuição das classes de resistência a carvão para grupos do ensaio de 2017 (cana planta), em que R: resistente, I: intermediário e S: suscetível	60
4.16	Atribuição das classes de resistência a carvão para grupos do ensaio de 2017 (cana soca), em que R: resistente, I: intermediário e S: suscetível	61
4.17	Atribuição das classes de resistência a mosaico para grupos do ensaio de 2016 (cana planta), em que R: resistente, I: intermediário e S: suscetível	62
4.18	Atribuição das classes de resistência a mosaico para grupos do ensaio de 2016 (cana soca), em que R: resistente, I: intermediário e S: suscetível	63
4.19	Atribuição das classes de resistência a mosaico para grupos do ensaio de 2017 (cana planta), em que R: resistente, I: intermediário e S: suscetível	64
4.20	Atribuição das classes de resistência a mosaico para grupos do ensaio de 2017 (cana soca), em que R: resistente, I: intermediário e S: suscetível	65

A.1 Média e desvio padrão das variedades desconsideradas nos modelos para incidência de carvão - ensaio 2016	73
A.2 Média e desvio padrão das variedades desconsideradas nos modelos para incidência de mosaico - ensaio 2016	74
A.3 Média e desvio padrão das variedades desconsideradas nos modelos para incidência de carvão - ensaio 2017	75
A.4 Média e desvio padrão das variedades desconsideradas nos modelos para incidência de mosaico - ensaio 2017	76
A.5 Estimativas dos parâmetros do modelo ZIBI para ensaio 2016 - planta (carvão))	82
A.6 Estimativas dos parâmetros do modelo ZIBB para ensaio 2016 - soca (carvão)	82
A.7 Estimativas dos parâmetros do modelo Binomial para ensaio 2017 - planta (carvão)	83
A.8 Estimativas dos parâmetros do modelo BB para ensaio 2017 - soca (carvão)	83
A.9 Estimativas dos parâmetros do modelo BB para ensaio 2016 - planta (mosaico)	83
A.10 Estimativas dos parâmetros do modelo BB para ensaio 2016 - soca (mosaico)	84
A.11 Estimativas dos parâmetros do modelo Binomial para ensaio 2017 - planta (mosaico)	84
A.12 Agrupamento final para conjunto de dados de carvão para ensaio 2016 (cana planta)	84
A.13 Agrupamento final para conjunto de dados de carvão para ensaio 2016 (cana soca)	85
A.14 Agrupamento final para conjunto de dados de mosaico para ensaio 2016 (cana planta)	86
A.15 Agrupamento final para conjunto de dados de mosaico para ensaio 2016 (cana soca)	86
A.16 Agrupamento final para conjunto de dados de carvão para ensaio 2017 (cana planta)	87
A.17 Agrupamento final para conjunto de dados de carvão para ensaio 2017 (cana soca)	88
A.18 Agrupamento final para conjunto de dados de mosaico para ensaio 2017 (cana planta)	88
A.19 Agrupamento final para conjunto de dados de mosaico para ensaio 2017 (cana soca)	89

1 INTRODUÇÃO

A cana-de-açúcar é uma das principais culturas do mundo, cultivada em mais de 100 países, sendo que o Brasil é o maior produtor mundial dessa cultura, seguido da Índia (RURAL, 2018). Grande parte desse sucesso deve-se aos Programas de Melhoramento Genético que selecionam as melhores variedades baseados, principalmente, em características de produtividade e resistência a doenças.

As doenças de plantas constituem uma das principais causas que limitam os rendimentos das culturas e a qualidade dos produtos (BUENO, 2010). Até hoje, foram identificadas 216 doenças que afetam a cana-de-açúcar, sendo que cerca de 58 foram encontradas no Brasil (SANTIAGO e ROSSETO, 2017). O carvão, o mosaico, a escaldadura das folhas, o raquitismo das soqueiras e a ferrugem são as mais importantes, pois seus danos podem representar um fator limitante na produção de variedades e é por isso que são consideradas em todos os programas de melhoramento genético (BUENO, 2010).

Doenças de plantas são vários processos fisiológicos que as prejudicam de forma constante, causados por algum microrganismo, chamado de patógeno, ou algum fator ambiental, resultando em sintomas específicos (CARDOSO, 2017). Elas ocorrem devido à interação dos fatores: ambiente, patógeno e hospedeiro (a planta), formando o triângulo de doença da Figura 1.1.

Figura 1.1. Triângulo da doença (CARDOSO, 2017)

Segundo CARDOSO (2017), se um desses vértices do triângulo é controlado, a planta não terá doenças, ou seja, para haver uma doença é necessário que o ambiente seja favorável à sobrevivência do patógeno, que o hospedeiro seja suscetível a esse patógeno e que exista um patógeno na área capaz de causar a doença. Dessa forma, seus efeitos podem variar de um local para outro, e somente com estudos e pesquisas para cada condição, é que se pode chegar ao estabelecimento da sua importância e aos meios mais eficientes de controle (SANGUINO, 2012).

1.1 Carvão

O carvão é uma doença da cana-de-açúcar causada pelo fungo *Sporisorium scitamineum*. Segundo SANTIAGO e ROSSETO (2017), os danos causados pelo fungo incidem tanto na redução da produção como na perda de qualidade do caldo. A doença caracteriza-se pelo surgimento de um chicote (Figura 1.2), que consiste em uma modificação da região de crescimento do colmo, induzida pelo fungo. O chicote é composto por parte do tecido da planta e parte do tecido do fungo. Inicialmente, esse chicote apresenta cor prateada, passando, posteriormente, à preta, devido à maturação dos esporos nele contidos. Antes de emitirem o chicote, as plantas doentes apresentam folhas estreitas e curtas, colmos mais finos que o normal e touceiras com superbrotaamento. Os chicotes surgem em plantas com idade entre dois e quatro meses, sendo que o pico ocorre entre seis e sete meses de idade.

Condições de estresse hídrico e calor favorecem a ocorrência do fungo e em casos extremos até variedades resistentes ao fungo podem apresentar sintomas da doença. A transmissão da doença se dá de forma aérea, por disseminação a partir dos chicotes e por meio do plantio de mudas infectadas (SANTIAGO e ROSSETO, 2016). As gemas representam o ponto de entrada e infecção e assim possuem um papel crucial na interação patógeno-hospedeiro. Gemas dormentes têm uma maior resistência à infecção

Figura 1.2. Chicote de carvão em cana-de-açúcar

mas gemas em processo de brotação têm um acesso mais fácil, aumentando as chances de penetração (BUENO, 2010).

1.2 Mosaico

O mosaico é uma doença causada pelo vírus do mosaico da cana-de-açúcar. Os sintomas ocorrem basicamente nas folhas, podendo variar a sua intensidade de acordo com a resistência da variedade, locais onde estão cultivadas e linhagem do vírus. Os sintomas mais evidentes são de áreas foliares com diferentes intensidades de verde intenso, circundadas por um verde mais claro ou áreas cloróticas de coloração mais amarelada (CANAOSTE, 2013). A doença causa perda de produtividade e longevidade do canavial, já que os sintomas podem progredir e provocar até a necrose da área foliar, sendo mais evidentes na base das folhas e em plantas mais jovens de rápido crescimento; em plantas mais velhas podem desaparecer dependendo do grau de resistência da variedade.

Figura 1.3. Mosaico da cana-de-açúcar

O vírus do mosaico é disseminado, naturalmente, por meio de várias espécies de afídeos (pulgões), que ao pousarem e darem uma “picada” em uma planta (com o seu estilete contaminado com vírus) transmitem a doença. Vale ressaltar que algumas culturas como milho e sorgo são hospedeiras de pulgões que podem estar infectadas por vírus do gênero *Potyvirus*, agentes causais do mosaico (CANAOSTE, 2013).

1.3 Justificativa e importância do trabalho

No caso do carvão e mosaico, a observação da incidência da doença é feita pela proporção de plantas sintomáticas. Entretanto, a caracterização é dada por meio de nota de suscetibilidade com escala de 1 a 9, sendo que notas mais altas representam maior suscetibilidade (Tabela 1.1). Existem poucos trabalhos apresentando metodologias para transformação da incidência em notas, em que a maioria deles utiliza escala fixa para classificação, como a proposta por CHAO et al. (1989) (Tabela 1.2). O maior

problema em utilizar escala fixa de classificação é que quando a incidência de doença é mais baixa, a nota de uma variedade pode ser subestimada.

Tabela 1.1. Caracterização de variedades quanto à resistência a doenças, seguindo a escala de HUTCHINSON e DANIELS (1971)

Nota	Classificação detalhada	Classificação geral
1	Altamente resistente	Resistente (R)
2	Resistente	
3	Moderadamente resistente	
4	Resistente/intermediária	Intermediária (I)
5	Intermediária	
6	Intermediária/Suscetível	
7	Moderadamente suscetível	Suscetível (S)
8	Suscetível	
9	Altamente suscetível	

Tabela 1.2. Classificação de variedades seguindo a escala de CHAO et al. (1989)

Taxa de infecção (%)	Classificação
[0 ; 11,0]	Resistente (R)
(11,0 ; 22]	Intermediária (I)
(22,0 ; 100]	Suscetível (S)

Grande parte dos programas de melhoramento genético de cana-de-açúcar realiza experimentos, chamados de fitotestes, para selecionar e caracterizar as variedades quanto a sua resistência a doenças. Nesses ensaios, é imprescindível inserir, como tratamentos, variedades já conhecidas quanto à resistência à doença a ser testada. Tais variedades são utilizadas como padrões de referência (**padrões**) para validação do ensaio e, também, para auxiliar na classificação das variedades testes (**clones**).

HUTCHINSON e DANIELS (1971) sugerem que a classificação deve ser feita por meio de comparações dos clones com os padrões, utilizando análise de regressão polinomial. Nesse contexto, as notas de suscetibilidade (escala de 1 a 9) dos padrões, definidas previamente, entram no modelo para explicar a porcentagem de plantas sintomáticas. Após o ajuste do modelo, constroem-se intervalos de confiança para as notas dos padrões, em que os clones são distribuídos a partir das suas médias.

STRINGER et al. (2012) sugerem uma técnica baseada na análise de modelo linear misto, em que os padrões são separados em grupos com base no teste de diferença mínima significativa e os clones são colocados no grupo em que melhor se ajustam. Nessa técnica, há necessidade de transformação dos dados e a correlação entre o número de grupos estatísticos formados em relação ao número de notas históricas indica a qualidade da classificação.

Muitos dados, relacionados a doenças, são do tipo contagem, em que se obtém o número de indivíduos doentes. Devido a sua natureza, esses dados não seguem uma distribuição normal, sendo essa uma pressuposição estabelecida nos modelos lineares clássicos. O auxílio de técnicas como a transformação de dados possibilita o uso dos modelos clássicos mesmo para dados desse tipo, entretanto, a transformação é alvo de críticas por pesquisadores como Ahrens et al. (1990), além de nem sempre ser capaz de contornar problemas de não atendimento às pressuposições e ainda comprometer a interpretação do modelo. Na literatura, os modelos lineares generalizados (MLG) são uma alternativa interessante para essa ocasião e está sendo cada vez mais acessada por diversas áreas, por englobarem diferentes tipos de distribuições. É

importante ressaltar que, na literatura, não foi encontrado nenhum trabalho classificando os clones por meio de modelos lineares generalizados.

Dessa forma, o objetivo desse trabalho é apresentar uma metodologia de classificação de variedades de cana-de-açúcar quanto à suscetibilidade, para as doenças carvão e mosaico, utilizando modelos lineares generalizados ajustados para proporção de plantas sintomáticas. Dentro da classe de modelos lineares generalizados e extensões, foram abordados os modelos binomial; quasi-binomial e beta-binomial (adequação de dados com superdispersão); binomial inflacionado de zeros e beta-binomial inflacionados de zeros (adequação de dados com excesso de zeros).

1.4 Nomenclatura

Cana-de-açúcar é uma cultura semi-perene, cujo manejo envolve, geralmente, quatro a cinco anos para haver novamente o plantio. Assim, aproveita-se a rebrota da cana durante cerca de três a quatro anos pós plantio. A seguir, são apresentados os termos comumente usados e, além disso, a Figura 1.4 apresenta o ciclo da cana-de-açúcar.

- **cana planta:** planta após plantio em que ainda não foi realizada nenhuma colheita (corte);
- **cana soca:** planta que teve um corte;
- **cana ressoca:** planta que já teve dois cortes ou mais;
- **fitoteste:** ensaio específico para testar a resistência de plantas quanto a doenças de interesse;
- **clone:** variedade em teste de suscetibilidade;
- **padrão:** variedade comercial, utilizada como referência de suscetibilidade;
- **tolete:** a cana-de-açúcar é propagada por toletes (mudas), provenientes dos colmos (canas);
- **gemas:** as gemas estão presentes nos toletes e são constituídas de células e tecidos em estado de latência e grande poder de diferenciação (AUDE, 1993).

Toletes e gema da cana-de-açúcar estão representados na Figura 1.5.

Figura 1.4. Ciclo da cana

Figura 1.5. Representação de toletes de cana-de-açúcar à esquerda e gema à direita

2 REVISÃO BIBLIOGRÁFICA

2.1 Modelos lineares generalizados (MLG)

Os modelos lineares generalizados (MLGs) são uma generalização dos modelos lineares clássicos com ampliação da distribuição da variável resposta, desde que a mesma pertença à família exponencial de distribuições (BATISTA, 2015). Segundo CORDEIRO e DEMÉTRIO (2013), os MLGs podem ser usados quando se tem uma única variável aleatória Y e associado a ela um conjunto de variáveis explicativas $X_1, X_2, \dots, X_p$. Envolve três componentes:

- a) **Componente aleatório:** representado por um conjunto de variáveis aleatórias independentes $Y_1, Y_2, \dots, Y_n$, provenientes de uma mesma distribuição da família exponencial na forma canônica, ou seja, com função de probabilidade (variável discreta) ou função de densidade de probabilidade (variável contínua), descrita sob a forma:

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

em que:

$$E(Y_i) = \mu_i = b'(\theta_i) \text{ e } \text{Var}(Y_i) = \phi b''(\theta_i) = \phi V(\mu_i)$$

sendo $V(\mu_i)$ a função de variância que depende unicamente da média μ_i , $\phi > 0$, o parâmetro de dispersão e θ_i o parâmetro canônico.

- b) **Componente sistemático:** responsável por especificar as variáveis explicativas $\mathbf{X}$ ao preditor linear $\boldsymbol{\eta}$, já que as variáveis explicativas entram no modelo na forma de uma soma linear de seus efeitos, isto é,

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta},$$

sendo $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$ a matriz de covariáveis de dimensões $n \times p$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$, o vetor de parâmetros a serem estimados e $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^T$, o vetor de preditores lineares.

- c) **Função de ligação:** função que relaciona o componente aleatório ao componente sistemático, ou seja, relaciona a média ao preditor linear, isto é,

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

sendo $g(\mu_i)$ uma função monótona e diferenciável.

Portanto, na escolha do MLG é importante definir a distribuição da variável resposta, a matriz do modelo e a função de ligação.

O processo de ajuste de MLG pode ser dividido em quatro etapas: (i) formulação dos modelos; (ii) ajuste dos modelos; (iii) verificação da adequação dos modelos e (iv) inferência. Tais etapas são realizadas sequencialmente. Na análise de dados complexos, após a realização da etapa de inferência, pode-se voltar à etapa (i) e escolher outros modelos que melhor se adequam aos dados (CORDEIRO e DEMÉTRIO, 2013).

A etapa da formulação dos modelos compreende a escolha de opções para a distribuição de probabilidade da variável resposta, covariáveis (matriz modelo) e função de ligação. Quanto mais se conhece a respeito dos dados, mais assertiva é a escolha, o que explica a importância da realização de uma boa análise exploratória. A escolha de uma função de ligação compatível com a distribuição

proposta para os dados, deve resultar de considerações a priori, exame intensivo dos dados, facilidade de interpretação do modelo e, mais usualmente, uma mistura de tudo isso (CORDEIRO e DEMÉTRIO, 2013).

2.1.1 Estimação dos parâmetros

Segundo CORDEIRO e DEMÉTRIO (2013), a etapa de ajuste representa o processo de estimação dos parâmetros lineares dos modelos e de determinadas funções das estimativas desses parâmetros, que representam medidas de adequação dos valores estimados. Existem vários métodos que podem ser usados para estimar os parâmetros dos MLG, como o qui-quadrado mínimo, o bayesiano e o de máxima verossimilhança. O principal método é o de máxima verossimilhança, por ter propriedades ótimas, tais como, consistência e eficiência assintótica.

O logaritmo da função de verossimilhança de uma amostra aleatória, proveniente de uma distribuição pertencente à família exponencial (Equação 2.1), é:

$$l(\beta) = \sum_{i=1}^n \ln(f(y_i; \theta_i, \phi)) = \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{\phi} + \sum_{i=1}^n c(y_i, \phi). \quad (2.2)$$

Uma propriedade da família exponencial de distribuições é que seus elementos satisfazem a condições de regularidade suficientes para assegurar que o máximo global do logaritmo da função de verossimilhança $l(\beta)$, seja dado unicamente pela solução do sistema de equações da função escore (U_j):

$$U_j = \frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l_i}{\partial \beta_j} = 0$$

para todo j . Aplicando a regra da cadeia, a função escore será:

$$U_j = \sum_{i=1}^n \left[\frac{y_i - \mu_i}{\phi} \frac{1}{V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \right], \quad j = 1, \dots, p.$$

Em geral, as equações $U_j = 0, j = 1, 2, \dots, p$ não são lineares e têm que ser resolvidas numericamente por processos iterativos do tipo Newton-Raphson ou escore de Fisher (CORDEIRO e DEMÉTRIO, 2013).

2.1.2 Função desvio e estatística X^2 de Pearson

Segundo MCCULLAGH e NELDER (1989), o ajuste de um modelo a um conjunto de dados observados $\mathbf{y}$ pode ser olhado como uma maneira de se substituir $\mathbf{y}$ por um conjunto de valores estimados $\hat{\mu}$ para um modelo com um número de parâmetros relativamente pequeno. Assim, admitindo-se uma combinação satisfatória da distribuição da variável resposta e da função de ligação, o objetivo é determinar quantos termos são necessários na estrutura linear para uma descrição razoável do processo gerador dos dados (CORDEIRO e DEMÉTRIO, 2013).

Considerando n observações, a elas podem ser ajustados modelos contendo até n parâmetros, chamado de *modelo saturado* ou *modelo completo* (não informativo). Por outro lado, temos o modelo mais simples chamado *nulo*, que tem um único parâmetro, representado por um valor $\hat{\mu}$ comum a todos os dados (ajuste pobre). Além disso, existem dois outros modelos, não tão extremos: o *modelo minimal* que contém o menor número de termos necessários para o ajuste, e o *modelo maximal* que inclui o maior número possível de termos que pode ser considerado.

Como o objetivo é determinar um modelo parcimonioso, que tenha o menor número de parâmetros possível e que seja capaz de explicar bem os dados, em geral, trabalha-se com modelos encaixados e o conjunto de matrizes dos modelos pode, então, ser formado pela adição sucessiva de termos ao modelo mais simples até se chegar ao modelo mais completo (CORDEIRO e DEMÉTRIO, 2013). Qualquer

modelo com p parâmetros linearmente independentes, situado entre os modelos minimal e maximal, é chamado *modelo corrente* ou *modelo sob pesquisa*.

O problema é determinar a utilidade de um parâmetro extra no modelo corrente (sob pesquisa) ou, então, verificar a falta de ajuste induzida pela omissão dele. A fim de discriminar entre modelos, medidas de discrepância devem ser introduzidas para medir o ajuste de um modelo, que para os MLGs são a estatística de Pearson generalizada X^2 e a função desvio (*deviance*). Segundo CORDEIRO e DEMÉTRIO (2013), a *deviance* tem grande vantagem por ser aditiva para um conjunto de modelos encaixados e é expressa por:

$$S_p = 2(\hat{l}_n - \hat{l}_p), \quad (2.3)$$

sendo, $\hat{l}_n$ e $\hat{l}_p$ os máximos dos logaritmo da função de verossimilhança (Equação 2.2) para os modelos com n parâmetros (*saturado*) e p parâmetros (*corrente*), respectivamente. Obtém-se, então,

$$\hat{l}_n = \frac{1}{\phi} \sum_{i=1}^n \left\{ \left[y_i \tilde{\theta}_i - b(\tilde{\theta}_i) \right] + c(y_i; \phi) \right\}$$

e

$$\hat{l}_p = \frac{1}{\phi} \sum_{i=1}^n \left\{ \left[y_i \hat{\theta}_i - b(\hat{\theta}_i) \right] + c(y_i; \phi) \right\},$$

sendo $\tilde{\theta}_i$ e $\hat{\theta}_i$ as estimativas de máxima verossimilhança do parâmetro canônico sob os modelos saturado e corrente, respectivamente. Assim tem-se:

$$S_p = \frac{2}{\phi} \sum_{i=1}^n \left\{ y_i \left[\tilde{\theta}_i - \hat{\theta}_i \right] - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \right\} = \frac{1}{\phi} D_p$$

em que S_p é chamado de desvio escalonado e D_p , desvio. Pode-se escrever ainda que:

$$S_p = \frac{1}{\phi} \sum_{i=1}^n d_i^2,$$

em que d_i^2 é o componente do desvio e mede a diferença dos logaritmos das funções de verossimilhança observada e ajustada, para a observação i correspondente.

Assumindo-se que o modelo corrente é verdadeiro, para ϕ conhecido, tem-se que S_p é, aproximadamente, distribuído como χ_{n-p}^2 . Assim, quando $S_p = \frac{1}{\phi} D_p \leq \chi_{(n-p; \alpha)}^2$, pode-se considerar que existem evidências, a um nível de significância α , que o modelo proposto está bem ajustado aos dados.

A estatística de Pearson X^2 generalizada é dada por:

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

sendo $V(\hat{\mu}_i)$ a função de variância estimada do modelo sob pesquisa.

2.1.3 Análise do desvio e seleção de modelos

A análise do desvio (ANODEV) é uma generalização da análise de variância (ANOVA) para os MLG, visando obter, a partir de uma sequência de modelos encaixados, os efeitos de variáveis explicativas, fatores e suas interações. Diz-se que M_q e M_p , com q e p parâmetros, são modelos encaixados se $M_q \subset M_p$ ($q < p$), tendo os modelos a mesma distribuição e a mesma função de ligação. A estatística $D_q - D_p$ com $(p-q)$ graus de liberdade é explicada pelos termos que estão em M_p e não estão em M_q .

Os métodos de inferência nos MLG baseiam-se na teoria de máxima verossimilhança. Usualmente utiliza-se a estatística da razão de verossimilhanças para testar hipóteses relativas aos parâmetros

β 's que, sob a hipótese nula H_0 , converge para uma variável aleatória com distribuição $\chi^2_{(p-q)}$, para ϕ conhecido. Assim, tem-se, assintoticamente:

$$S_q - S_p = \phi^{-1}(D_q - D_p) \sim \chi^2_{(p-q)}.$$

Se ϕ é desconhecido, deve-se obter uma estimativa $\hat{\phi}$ consistente, de preferência baseada no modelo maximal (com m parâmetros), e a inferência pode ser baseada na estatística F, expressa por:

$$F = \frac{(D_q - D_p)/(p - q)}{\hat{\phi}} \sim F_{p-q, n-m}.$$

Assim, se $S_q - S_p > \chi^2_{q-p; (1-\alpha)}$ (ou $F > F_{q-p, n-m; (1-\alpha)}$) rejeita-se a hipótese nula de que o efeito do parâmetro adicional é zero, a um nível de significância de nível α , ou seja, os parâmetros testados são importantes para descrever os dados e devem permanecer no modelo.

Outras estatísticas utilizadas como medidas de comparação de qualidade de ajuste do modelo são os critérios de informação de Akaike (AIC) e de informação Bayesiano (BIC), propostos por Akaike (1973) e definidos respectivamente, por:

$$AIC = -2\hat{l}_p + 2p$$

e

$$BIC = -2\hat{l}_p + p \log n$$

em que p é o número de parâmetros do modelo sob pesquisa, $\hat{l}_p$ é o logaritmo da função de máxima verossimilhança para o modelo sob pesquisa, e n é o número de observações do conjunto de dados. A escolha do melhor modelo se faz considerando aquele que apresenta o menor valor dos critérios utilizados (AIC ou BIC).

2.1.4 Estimação do parâmetro de dispersão

Quando ϕ é desconhecido, admite-se que ele seja o mesmo para todas as observações, isto é, constante (CORDEIRO e DEMÉTRIO, 2013). Os métodos do desvio e de Pearson são os mais utilizados para sua estimação. O primeiro é definido por

$$\hat{\phi} = \frac{D_p}{n - p}, \quad (2.4)$$

e o segundo, apresentado por MCCULLAGH e NELDER (1989), é dado por:

$$\hat{\phi} = \frac{X^2}{n - p},$$

sendo que D_p é o desvio, X^2 é a estatística generalizada de Pearson e $(n - p)$ é o grau de liberdade do modelo maximal.

Para modelos em que ϕ é conhecido, os estimadores acima são muito utilizados para verificação da qualidade de ajuste de modelos, pois se o estimador diferir muito do parâmetro de dispersão teórico, tem-se indícios de não adequação da função de variância $V(\hat{\mu}_i)$, como, por exemplo, superdispersão que será abordada na Seção 2.2.2.

2.1.5 Técnicas para diagnósticos em MLG

As técnicas de diagnóstico são de grande importância para validação do modelo ajustado, permitindo identificar: não adequação da distribuição proposta; escolha inadequada da função de ligação e/ou função de variância; presença de pontos discrepantes devidos a pontos extremos, ou que foram erroneamente coletados ou foram gerados por algum motivo não controlado. Para os MLGs, as técnicas são semelhantes às usadas em modelos clássicos, porém, com algumas adaptações.

2.1.5.1 Tipos de resíduos

Na classe dos MLGs existem vários tipos de resíduos. Entretanto, são apresentados os resíduos ordinários, os componentes da *deviance* residual, os componentes da *deviance* residual estudentizada e o resíduo quantílico aleatorizado, conforme se segue.

(i) Resíduos ordinários:

$$r_i = y_i - \hat{\mu}_i,$$

em que y_i representa a variável resposta e $\hat{\mu}_i$, sua respectiva média estimada.

(ii) Componentes da *deviance* residual:

$$r_i = \text{sinal}(y_i - \hat{\mu}_i) \times \sqrt{d_i},$$

em que $d_i = 2 \left\{ y_i[\tilde{\theta} - \hat{\theta}] - b(\tilde{\theta}) + b(\hat{\theta}) \right\}$ e $\tilde{\theta}_i$ e $\hat{\theta}_i$ são as estimativas de máxima verossimilhança do parâmetro canônico sob os modelos saturado e corrente, respectivamente. Esse resíduo será a raiz quadrada de cada elemento do desvio com o sinal igual ao de $y_i - \hat{\mu}_i$.

(iii) Componentes da *deviance* residual estudentizada:

$$r_i = \frac{\text{sinal}(y_i - \hat{\mu}_i) \times \sqrt{d_i}}{\sqrt{\hat{\phi}(1 - \hat{h}_{ii})}},$$

em que $\hat{\phi}$ é uma estimativa do parâmetro de dispersão ϕ , e $\hat{h}_{ii}$ são os elementos da diagonal da matriz de projeção $\mathbf{H}$, expressa por

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

sendo $\mathbf{X}$ a matriz do modelo e $\mathbf{W}$, de pesos estimados. A distribuição r_i nesse caso está mais próxima da normalidade.

(iv) Resíduo quantílico aleatorizado:

$$r_i = \Phi^{-1} \left\{ F(y_i; \hat{\mu}_i, \hat{\phi}) \right\}, \quad (2.5)$$

em que $\Phi(\cdot)$ é a acumulada da função de distribuição normal padrão e $F(y_i; \mu_i, \phi)$ é a acumulada da distribuição $f(y_i, \mu_i, \phi)$ (DUNN e SMYTH, 1996).

2.1.5.2 Técnicas gráficas

Entre as técnicas de análise dos resíduos destacam-se:

- (i) Gráfico de resíduos versus a ordem das observações para avaliar independência dos dados (quando a ordem das observações tem algum significado prático). Nesse caso, os pontos devem estar distribuídos aleatoriamente (sem uma tendência definida);
- (ii) Gráfico de resíduos versus as variáveis explicativas ou versus valores ajustados para avaliar a função de variância. Considera-se que a função de variância está adequada se os pontos estiverem distribuídos aleatoriamente no gráfico (sem uma tendência definida);
- (iii) Gráfico entre a variável dependente ajustada e o preditor linear, que segundo CORDEIRO e LIMA NETO (2006) se a tendência dos dados plotados no gráfico for aproximadamente linear, a função de ligação estará correta. Entretanto, para dados binários este gráfico é não informativo, sendo necessário o uso de métodos formais.
- (iv) Gráfico semi-normal de probabilidades com envelope simulado (*half normal plot - hnp*), que permite averiguar a adequação da distribuição dos erros e função de ligação, e, também, detectar superdispersão e *outliers*. Segundo MORAL et al. (2017), essa técnica consiste em plotar os valores absolutos ordenados de uma medida de diagnóstico (resíduo) versus os “*half-normal scores*” (estatísticas de ordem esperadas), que pode ser aproximada por:

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

em que i é a i -ésima estatística de ordem, $1 \leq i \leq n$, n é o tamanho da amostra e Φ é a função acumulada da distribuição normal padrão.

A obtenção do envelope simulado para um gráfico semi-normal consiste em:

- a) ajustar o modelo;
- b) extrair um diagnóstico do modelo (resíduos) e ordenar os valores absolutos;
- c) simular 99 (ou mais) amostras para a variável resposta usando a mesma matriz do modelo, a mesma distribuição de erros e mesmas estimativas de parâmetros;
- d) ajustar o mesmo modelo para cada amostra simulada e extrair o mesmo diagnóstico do modelo e, novamente, ordenar os valores absolutos;
- e) calcular os percentis desejados dos valores dos diagnósticos simulados (que irão compor os envelopes do gráfico);
- f) plotar os percentis obtidos versus os “*half-normal scores*”.

MORAL (2013) implementou a função *hnp()* no pacote *hnp()* do R, facilitando o uso da técnica para diversas funções de probabilidade, diferentes funções de ligação e para diferentes tipos de resíduos. Se a maior parte dos pontos estiverem contidos no envelope simulado, há indícios de bom ajuste do modelo.

2.1.5.3 Verificação da função de ligação

Um dos procedimentos formais segundo CORDEIRO e LIMA NETO (2006) é o método proposto por Hinkley (1985), que consiste em adicionar $\hat{\eta}^2$ como uma nova covariável na matriz modelo. Se isto causar uma redução significativa no desvio, a função de ligação não é adequada. Para verificar se a redução é estatisticamente significativa, pode-se utilizar a análise do desvio (ANODEV).

2.2 Modelos para dados de proporção

2.2.1 Modelo binomial

Considerado, inicialmente, para estudo de dados na forma de proporções, em que se têm diferentes tratamentos aplicados a n conjuntos de indivíduos, sendo m_i o número de indivíduos testados no conjunto i , $i = 1, \dots, n$. Assume-se que os indivíduos são independentes e possuem a mesma probabilidade de sucesso π_i . Seja Y_i uma variável aleatória que representa o número de sucessos em uma amostra de tamanho m_i . É razoável assumir que $Y_i \sim \text{Bin}(m_i, \pi_i)$. Dessa forma, tem-se que:

$$P(Y_i = y_i) = \binom{m_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{m_i - y_i},$$

com

$$E[Y_i] = \mu_i = m_i \pi_i$$

e

$$\text{Var}[Y_i] = m_i \pi_i (1 - \pi_i),$$

em que $y_i = 0, 1, \dots, m_i$ e $0 < \pi_i < 1$.

A função de ligação canônica para o modelo binomial é a logística (*logit*):

$$g(\mu_i) = \log \left(\frac{\pi_i}{1 - \pi_i} \right).$$

Essa função de ligação mapeia as probabilidades π_i para a linha real e produz uma curva com característica sigmoidal, usada para descrever o comportamento da variável em experimentos de dose-resposta (FATORETTO et al., 2018), como representado na Figura 2.1.

Figura 2.1. Representação da função de ligação *logit*

No caso do modelo binomial, tem-se $\phi = 1$. Sendo assim, como visto na Seção 2.1.4, o estimador $\hat{\phi}$ pode ser utilizado para verificação de adequação da função de variância do modelo, que pode sinalizar indícios de superdispersão ($\hat{\phi} > 1$) ou subdispersão ($\hat{\phi} < 1$).

2.2.2 Superdispersão

Um conjunto de dados pode apresentar variabilidade (*deviance* residual) maior do que a esperada pelos modelos probabilísticos padrões. Esse fenômeno é conhecido como superdispersão e tem sido amplamente considerado na literatura, principalmente em relação às distribuições binomial e de poisson.

Um importante fato sobre a distribuição de probabilidade da binomial é que sua variância tem direta relação com a média μ_i , como visto na Figura 2.2. No entanto, quando a superdispersão está presente, a variância será maior do que isso, ou seja, **não** acompanhará o comportamento esperado.

Figura 2.2. Distribuição da variância da distribuição de probabilidade binomial

Existem várias possíveis causas que podem gerar dados com superdispersão, como a má-especificação do preditor linear, presença de *outliers*, função de ligação inadequada, proporções com denominador reduzido, dependência longitudinal não ajustada e ocorrência de agrupamentos (clusters) que tornam a probabilidade de sucesso π não constante para todas as observações (DEMÉTRIO e HINDE, 1997). Uma fonte particular de superdispersão é quando proporções ou contagens são inflacionadas de zeros.

Para dados com excesso de zeros, é razoável a suposição de haver mais de um processo gerador atuando na geração dos dados (RIDOUT et al., 1998). Assim, a ocorrência de valores nulos pode ser caracterizada como:

- zeros amostrais: Ocorrem segundo um processo gerador de contagens;
- zeros estruturais: Ausência de determinada característica da população.

No caso de doenças em plantas, pode-se pensar em ambos os casos:

- zeros amostrais: Quando algum vértice do triângulo da doença (Figura 1.1) não está favorável para apresentação de sintomas na planta;
- zeros estruturais: A variedade é resistente à doença.

Segundo DEMÉTRIO (1994), na prática, a superdispersão é mais comum do que a dispersão dada pela distribuição sob consideração. O grau e a incidência com que ela aparece depende muito do campo de aplicação. Em estudos de incidências de doenças, em que se trabalha com proporções, de forma geral, a variância do modelo binomial é insignificante em relação à variância total. A superdispersão ocorre naturalmente quando os dados são provenientes de famílias, ninhadas, colônias, e pessoas que habitam a mesma casa.

Além disso, existem outros fatores que podem causar variação no conjunto de dados, como a dose aplicada de inoculação, que está sujeita a erros de medição (em termos de potencial de inóculo) e a seleção de isolados do fungo, pois ao fazer um ensaio para esse fim (fitoteste), a massa de esporos usados, em geral, vem do campo, e é composta de indivíduos suscetíveis, intermediários e resistentes, gerando uma mistura de distribuições.

A superdispersão é bastante recorrente em estudos biológicos e deve ser tratada adequadamente na análise, uma vez que, ignorá-la pode resultar em inferência enganosa para os parâmetros de regressão, já que a matriz de variâncias e covariâncias das estimativas dos parâmetros fica subestimada (DEMÉTRIO e HINDE, 1997). Por esse motivo, existem na literatura diversos modelos que incorporam a superdispersão, possibilitando melhores ajustes para esses tipo de dados. Pode-se categorizar os modelos em grupos que seguem duas diferentes abordagens:

- (i) Assumir forma mais geral para a função de variância, possivelmente, incluindo parâmetros adicionais. Modelos podem ser vistos como extensões úteis do modelo base;
- (ii) Supor modelo de dois estágios para a variável resposta, ou seja, que o próprio parâmetro do modelo base tenha alguma distribuição, resultando em um modelo de probabilidade composta para a variável resposta.

Ainda, dentro da segunda categoria, para dados com excesso de zeros existem duas principais abordagens como modelos alternativos:

- Modelos de barreira (*Hurdle Models*): que desconsideram os zeros amostrais e modelam, separadamente, os zeros estruturais e as contagens positivas;
- Modelos de mistura (*Zero Inflated Models*): que modelam os zeros (estruturais e amostrais) em conjunto com as contagens positivas.

Nas próximas seções serão apresentados alguns modelos especiais, que podem ser utilizados como alternativa para ajuste de modelos com superdispersão e excesso de zeros. Tem-se o modelo quasi-binomial da categoria (i) e os modelos beta-binomial, binomial inflacionado de zeros e beta-binomial inflacionado de zeros da categoria (ii).

2.2.3 Modelo quase-binomial

WEDDERBURN (1974) define a teoria de quase-verossimilhança, que vem generalizar a teoria de modelos lineares generalizados, permitindo que a suposição do conhecimento da distribuição seja substituída por uma pressuposição mais fraca, isto é, de que apenas o primeiro e segundo momentos da variável resposta sejam conhecidos. Assim, foi proposta uma função de estimação que, sob hipóteses gerais, leva a estimadores consistentes e assintoticamente normais dos parâmetros do modelo.

Segundo PAULA (2013), se Y_i é a variável aleatória de interesse, o logaritmo da função quase-verossimilhança é definido como:

$$Q = Q(y; \mu) = \frac{1}{\phi} \int_y^\mu (y - t)V(t)^{-1} dt, \quad (2.6)$$

em que $V(t)$ é uma função positiva conhecida e $\phi > 0$ é o parâmetro de dispersão.

Desse modo, seja Y_i uma variável aleatória que representa o número de sucessos em uma amostra de tamanho m_i , assumindo que $V(t) = t(1-t)$, $0 \leq y \leq 1$, $0 \leq t \leq 1$ e integrando (2.6) obtém-se a função de quase-verossimilhança e conclui-se que, nesse caso, o modelo de quase-verossimilhança é equivalente ao modelo binomial. Além disso, Q é proporcional ao logaritmo da função de verossimilhança do modelo binomial quando $\phi = 1$, para a função de variância assumida.

Assim, tem-se para o modelo quase-binomial que:

$$E[Y_i] = \mu_i = m_i \pi_i$$

e

$$\text{Var}[Y_i] = \phi m_i \pi_i (1 - \pi_i),$$

em que ϕ é o parâmetro de dispersão desconhecido e pode ser estimado como apresentado na Seção 2.1.4.

2.2.4 Modelo beta-binomial

Uma das premissas do modelo binomial é que a probabilidade de sucesso é a mesma para todos os indivíduos. Uma alternativa é permitir que a probabilidade de sucesso varie de acordo com alguma distribuição. Sendo assim, a distribuição beta-binomial resulta de uma mistura da distribuição binomial e a distribuição beta.

Seja Y_i uma variável aleatória que representa o número de sucessos em uma amostra de tamanho m_i e assumindo que $Y_i \sim \text{Bin}(m_i, \pi_i)$, com $\pi_i \sim \text{Beta}(\alpha_i, \beta_i)$. Tem-se marginalmente que Y_i tem uma distribuição beta-binomial com média e variância dadas, respectivamente, por:

$$P(Y_i = y_i) = \binom{m_i}{y_i} \frac{B(\alpha_i + y_i, \beta_i + m_i - y_i)}{B(\alpha_i, \beta_i)},$$

$$E[Y_i] = \mu_i = m_i \pi_i,$$

e

$$\text{Var}[Y_i] = m_i \pi_i (1 - \pi_i) [1 + \phi(m_i - 1)],$$

para $y_i = 0, 1, \dots, m_i$ e $0 < \pi_i < 1$, sendo que $\pi_i = \frac{\alpha_i}{\alpha_i + \beta_i}$ e $\phi = \frac{1}{\alpha_i + \beta_i + 1}$, com $\alpha_i + \beta_i$ constante. Nota-se, nesse caso, que para $\phi = 0$, tem-se o modelo binomial.

2.2.5 Modelos inflacionados de zeros

Os modelos inflacionados de zeros têm sido muito utilizados para modelagem de dados com excesso de zeros, em diferentes áreas de aplicação. É assumido que com probabilidade p a única observação possível é 0, e com probabilidade $1 - p$ é observada uma variável aleatória que descreve contagens de defeitos no estado imperfeito (estado em que o número de defeitos do produto segue uma distribuição de probabilidade) (MONZÓON MONTTOYA, 2009). Por exemplo, quando o ambiente não é favorável para doença (estado perfeito para planta), pode não haver planta doente (defeituosa). Senão, as plantas podem apresentar sintomas de acordo com uma distribuição do estado imperfeito.

A forma geral da função de massa de probabilidade de uma variável aleatória inflacionada de zeros discreta não negativa, é dada por:

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)f(y_i, \mu) & , y_i = 0 \\ (1 - p_i)f(y_i, \mu) & , y_i = 1, 2, \dots \end{cases}$$

em que $p_i \in [0, 1]$ e $\mu \geq 0$.

Para dados de proporção, dentro dessa categoria têm-se, dentre outros, os modelos binomial inflacionado de zeros (ZIB) e o beta-binomial inflacionado de zeros (ZIBB), que são alternativas para controlar a superdispersão.

2.2.5.1 Modelo binomial inflacionado de zeros - ZIBI

O modelo de regressão binomial com excesso de zeros tem a seguinte forma:

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)(1 - \pi_i)^{m_i} & , y_i = 0 \\ (1 - p_i)\binom{m_i}{y_i}\pi_i^{y_i}(1 - \pi_i)^{m_i - y_i} & , y_i = 1, 2, \dots, m_i, \end{cases}$$

em que $0 \leq p_i < 1$ e $0 < \pi_i < 1$, sendo $\alpha_i + \beta_i$ constante. Seus primeiros momentos são, respectivamente, dados por:

$$E[Y_i] = (1 - p_i)m_i\pi_i$$

e

$$\text{Var}[Y_i] = (1 - p_i)m_i\pi_i[1 - \pi_i(1 - p_im_i)].$$

Os parâmetros p_i e μ_i são modelados, usualmente, com funções de ligação $\log\left(\frac{p_i}{1-p_i}\right)$ e $\log\left(\frac{\pi_i}{1-\pi_i}\right)$. Neste caso, $\phi = 1$. Nota-se que para $p = 0$, ou seja, a probabilidade de $Y = 0$ é nula, tem-se o modelo binomial apresentado na Seção 2.2.1.

2.2.5.2 Modelo beta-binomial inflacionado de zeros - ZIBB

Seja $Y = 0$ com probabilidade p e $Y \sim$ beta-binomial com probabilidade $(1 - p)$, pode-se apresentar uma formulação do modelo beta-binomial com excesso de zeros por:

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)\frac{B(\alpha_i, \beta_i + m_i)}{B(\alpha_i, \beta_i)} & , y_i = 0 \\ (1 - p_i)\binom{m_i}{y_i}\frac{B(\alpha_i + y_i, \beta_i + m_i - y_i)}{B(\alpha_i, \beta_i)} & , y_i = 1, 2, \dots, m_i \end{cases}$$

em que $0 \leq p_i < 1$, $0 < \pi_i < 1$, $\pi_i = \frac{\alpha_i}{\alpha_i + \beta_i}$ e $\phi = \frac{1}{\alpha_i + \beta_i + 1}$ e seus primeiros momentos são, respectivamente, dados por:

$$E[Y_i] = (1 - p_i)m_i\pi_i$$

e

$$\text{Var}[Y_i] = (1 - p_i)m_i\pi_i(1 - \pi_i)[1 + \phi(m_i - 1)] + p_i(1 - p_i)m_i^2\pi_i^2,$$

Nota-se, nesse caso, que para $p = 0$, ou seja, a probabilidade de $Y = 0$ é nula, tem-se o modelo beta-binomial apresentado na Seção 2.2.4.

2.3 Pacote GAMLSS

STASINOPOULOS e RIGBY (2007) implementaram o pacote GAMLSS no *software* R, que permite o ajuste de modelos aditivos generalizados para localização, escala e forma (*generalized additive models for location, scale and shape* – GAMLSS). No pacote foram implementados modelos com várias distribuições de probabilidade e foram incluídas tanto distribuições simples, que pertencem à família exponencial, quanto várias distribuições que envolvem três ou quatro parâmetros.

Na Tabela 2.1, estão apresentadas as parametrizações utilizadas pelo pacote para os modelos binomial (BI), beta-binomial (BB), binomial inflacionado de zeros (ZIBI) e beta-binomial inflacionado de zeros (ZIBB), conforme apresentado em RIGBY et al. (2019).

Tabela 2.1. Parametrização dos modelos no pacote GAMLSS

Distribuição	ϕ	$E[Y]$	$\text{Var}[Y]$
BI $\sim(m, \mu)$	1	$m\mu$	$m\mu(1 - \mu)$
BB $\sim(m, \mu, \sigma)$	$\frac{\sigma}{(1+\sigma)}$	$m\mu$	$m\mu(1 - \mu)[1 + \phi(m - 1)]$
ZIBI $\sim(m, \mu, \sigma)$	1	$(1 - \sigma)m\mu$	$m\mu(1 - \sigma)[1 - \mu + m\mu\sigma]$
ZIBB $\sim(m, \mu, \sigma, \tau)$	$\frac{\sigma}{(1+\sigma)}$	$(1 - \tau)m\mu$	$(1 - \tau)m\mu(1 - \mu)[1 + \phi(m - 1)] + \tau(1 - \tau)m^2\mu^2$

Além disso, o pacote disponibiliza o desvio global D_g , dado por:

$$D_g = -2 * \hat{l}_p,$$

sendo $\hat{l}_p$ o máximo do logaritmo da função de verossimilhança para o modelo com p parâmetros (corrente).

O desvio global D_g , diferente da *deviance* S_p (Equação 2.3), pode ser utilizado para comparar modelos com diferentes distribuições (desde que os modelos sejam encaixados), já que inclui todos os termos constantes no logaritmo da função de verossimilhança (RIGBY et al., 2015).

3 MATERIAL E MÉTODOS

3.1 Material

Neste trabalho, foram analisados dados provenientes de dois experimentos de caracterização e seleção de clones quanto à suscetibilidade (**fitotestes**) a carvão e mosaico, realizados em 2016 e 2017, na cidade de Piracicaba, SP. Sabe-se que as duas doenças (carvão e mosaico) atacam as plantas de forma independente, de maneira que a avaliação dos sintomas visuais de carvão não atrapalha no mosaico, e vice-versa. Assim, as doenças foram avaliadas no mesmo ensaio, economizando-se mudas e espaço em campo.

3.1.1 Preparação das mudas e inoculação de carvão

Os ensaios se iniciaram pelo plantio dos toletes em bandejas para preparação das mudas, seguindo as seguintes etapas:

- a) Foram montadas cinco bandejas para cada tratamento (clones e padrões) para compor as cinco repetições do campo. Cada repetição foi colocada em uma bancada diferente dentro da casa de vegetação;
- b) Os toletes foram irrigados por aspersão, de maneira uniforme com sistema de irrigação;
- c) Inocularam-se, manualmente, com bombas de aspersão, os esporos de carvão em todas as gemas expostas. Foi utilizado o mesmo caldo de inóculo em cada bancada;
- d) Cobriram-se os toletes com substrato.

As bandejas ficaram, aproximadamente, 45 dias em casa de vegetação e foram levadas ao pátio para rustificação, ou seja, a adaptabilidade da muda às condições de pleno sol encontrada no campo. Após 30 dias, foram levadas a campo.

Figura 3.1. 1) Distribuição dos toletes nas bandejas 2) Inoculação do carvão 3) Bandejas inoculadas, cobertas com substrato 4) Plantas após 30 dias.

3.1.2 Linha infectora do mosaico

Na instalação do ensaio em campo, foi plantada uma linha infectora a cada quatro linhas de tratamentos, conforme descrito na Figura 3.2. Essa linha infectora foi composta por sorgo, milho doce e canas suscetíveis ao mosaico.

Cerca de três meses após a instalação do ensaio, foram coletadas, no campo, canas de diferentes variedades com presença de sintomas característicos da doença mosaico para a preparação de um líquido contaminante. As folhas coletadas foram picadas e batidas em um liquidificador com solução tampão. O líquido foi pulverizado, manualmente, nas linhas infectoras.

Foi verificada a presença do pulgão *Rhopalosiphum maidis* que é vetor do mosaico, assegurando que os vírus inoculados nas linhas infectoras seriam levados para os clones e padrões do ensaio.

Figura 3.2. Esquema da linha infectora de mosaico nos ensaios de campo

3.1.3 Experimentos de campo

Os ensaios foram montados no delineamento casualizado em blocos com 5 repetições. As parcelas, que compunham as unidades experimentais, foram compostas por duas linhas de cinco metros, totalizando 20 plantas cada (mudas excedentes foram descartadas), conforme Figura 3.2.

Como tratamentos, têm-se clones e padrões de carvão e mosaico, totalizando 104 tratamentos para o ensaio de 2016, e 54 para o ensaio de 2017, em que apenas os padrões de referência são comuns entre eles (Tabela 3.1). Os padrões possuem classificação histórica quanto à suscetibilidade a carvão e mosaico, conforme descrito na Tabela 3.2, entretanto, o padrão REF_10 não foi inserido no ensaio de 2016.

Tabela 3.1. Descrição dos ensaios

	Plantio 2016	Plantio 2017
Clones	95	44
Padrões	9	10
Tratamentos	104	54
Blocos	5	5
Parcelas	520	270

Tabela 3.2. Classificação histórica dos padrões

		Carvão		
		Resistente	Intermediário	Suscetível
Mosaico	Resistente	REF_9	REF_8	-
	Intermediário	REF_4 REF_5	-	-
	Suscetível	-	REF_1 REF_3	REF_2 REF_7 REF_6 REF_10

Foram feitas duas avaliações em cana planta e duas em cana soca. Essa prática é utilizada pois existe o tempo de aparecimento dos sintomas (latência) que pode variar em função das condições climáticas, além de garantir pelo menos uma avaliação por estágio de corte, em caso de tombamento. Entretanto, para fins de análise, utiliza-se a incidência máxima observada em cada corte.

As variáveis estudadas foram:

- proporção de touceiras com sintomas visuais de carvão (chicote);
- proporção de touceiras com sintomas visuais de mosaico.

3.2 Métodos

3.2.1 Análises exploratórias

Primeiro, avaliou-se a estrutura dos dados, com o intuito de detectar qualquer problema de delineamento. Em ensaios de campo, é natural que algumas plantas morram, causando falhas nas parcelas, ou seja, uma redução do número de unidades experimentais estabelecidas. Se houve muitas falhas, as estimativas dos parâmetros dos modelos poderão ser muito fracas, gerando resultados inconclusivos ou viesados.

Identificou-se a presença de variedades com variância nula, ou seja, que não apresentaram nenhuma planta sintomática para determinada doença ou, por outro lado, todas as plantas foram infectadas pela doença. Tais tratamentos não trazem nenhuma contribuição para o ajuste do modelo, além de causarem problemas de convergência e por isso seus dados foram desconsiderados nos ajustes dos modelos.

Realizou-se uma análise exploratória para avaliar o comportamento dos padrões de referência para cada uma das doenças nos ensaios e, posteriormente, realizou-se uma sequência de análises gráficas para identificar e avaliar características importantes dos dados:

- comportamento da distribuição dos dados e de valores nulos (parcelas com 0 plantas infectadas), facilitando a identificação de modelos potenciais;
- distribuição da incidência de doença por variedade em cada conjunto de dados, facilitando perceber possíveis problemas de superdispersão;
- comparação, correlação e interação entre os dados de cana planta e soca.

3.2.2 Modelagem

Neste estudo, a probabilidade de sucesso é de a planta ser sintomática para a doença em teste. Assim, Y_{ij} é o número de plantas sintomáticas, de um total de m_{ij} plantas avaliadas. Foram analisados, isoladamente, os dados de cana planta e cana soca para cada ensaio. Inicialmente, admitiu-se

a distribuição de probabilidade binomial, ou seja, $Y_{ij} \sim \text{BI}(m_{ij}, \pi_{ij})$. A função de ligação utilizada foi a *logit* e preditor linear utilizado foi:

$$\eta_{ij} = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \alpha + \gamma_i + \beta_j \quad (3.1)$$

em que $i = 1, 2, \dots, 5$ e $j = 1, 2, \dots, t$, α é a constante, γ_i é o efeito do i -ésimo bloco e β_j o efeito do j -ésimo tratamento. A qualidade de ajuste dos modelos foi avaliada pelo gráfico semi-normal de probabilidades com envelope de simulação (*hnp*) e teste de razão de verossimilhanças.

Como cada observação (parcela) possui no máximo 20 plantas avaliadas, acredita-se que quando apenas uma planta, de apenas uma repetição, apresenta sintoma, a variância para esse tratamento poderá ser inflacionada, uma vez que temos apenas cinco repetições por tratamento. Assim, para os dados cujos modelos que tiveram indicativo de bom ajuste pelo teste de razão de verossimilhanças e parâmetros de dispersão estimados muito próximos do teórico ($\phi = 1$), mas apresentaram muitos pontos próximos de zero fora do envelope (*hnp*), uniu-se as variedades cujos dados possuíam esta configuração e ajustou-se o modelo binomial novamente.

Problemas de superdispersão foram avaliados, de forma exploratória, por meio das estimativas do parâmetro de dispersão (Equação 2.4). Para todos os conjuntos de dados, ajustou-se, também, os modelos quase-binomial (QBI), beta-binomial (BB), binomial inflacionado de zeros (ZIBI) e beta-binomial inflacionado de zeros (ZIBB), com funções de ligação descritas na Tabela 3.3. Dentre os modelos que tiveram indícios de bom ajuste pelo *hnp*, selecionou-se aquele com menor valor de AIC. Também, observaram-se as estimativas de BIC e desvio global para escolha do melhor modelo.

Existem vários fatores que podem causar superdispersão em variáveis dos conjuntos de dados avaliados, destacando-se as seguintes situações:

- dose aplicada de inoculação de carvão, que está sujeita a erros de medição (em termos de potencial de inóculo) e a seleção de isolados do fungo, como mencionado na Seção 2.2.2. O mesmo pode ocorrer para o vírus do mosaico nas linhas infectoras;
- aninhamento de pulgões (transmissores) nas linhas infectoras de mosaico, gerando uma inoculação não homogênea;
- ensaios realizados em campo, estando expostos a inoculações naturais das doenças, ou seja, não há controle da inoculação de forma que ela pode ocorrer em diferentes pressões nas parcelas.

Além disso, como o carvão é inoculado na gema, como visto na Seção 3.1.1, qualquer falha na inoculação acarretará na falha do vértice “patógeno” do triângulo da doença (Figura 1.1), aumentando a probabilidade de a planta não apresentar sintomas e consequentemente gerar um excesso de zeros no conjunto de dados.

Outras técnicas gráficas para análise da qualidade do ajuste dos modelos foram utilizadas: a) boxplot dos resíduos por variedade, para detecção de possíveis outliers; b) gráfico de correlação entre os resíduos e a ordem das parcelas, para detecção da presença (ou ausência) de correlação entre os dados, que poderia levar à necessidade de modelos que consigam capturar tal correlação e c) gráfico de correlação entre os resíduos e valores preditos para avaliar a adequação da função de variâncias. Espera-se, para um bom ajuste, que os resíduos estejam distribuídos aleatoriamente em torno de zero.

Todas as análises foram realizadas no software R. Para ajuste dos modelos utilizou-se a função *glm* para os modelos binomial e quase-binomial, e os demais modelos foram ajustados pelo pacote *gamlss* (RIGBY et al., 2019). Os gráficos semi-normais foram construídos, usando-se partir do pacote *hnp*.

Tabela 3.3. Descrição das funções de ligação utilizadas nos modelos, seguindo a parametrização do pacote *gamlss*

Distribuições	μ	σ	τ
BI	logit	-	-
BB	logit	log	-
ZIBI	logit	logit	-
ZIBB	logit	log	logit

A fim de entender os ganhos no uso de modelos mais elaborados no ajuste dos dados, tomaram-se os parâmetros estimados pelos modelos ajustados e, conforme descritos na tabela 2.1 na Seção 2.3, estimaram-se suas esperanças e variâncias. Então, avaliaram-se graficamente as funções de variância para modelos binomiais em comparação com os modelos finais de cada conjunto de dados.

3.2.3 Agrupamento de variedades

Após a escolha do modelo final, o interesse foi em agrupar variedades com comportamento semelhante, de acordo com as estimativas do modelo. Tal agrupamento foi realizado a partir do teste de razão de verossimilhanças para modelos encaixados, sequencialmente, de forma independente para cada conjunto de dados. Dessa forma, agrupou-se o maior número de variedades, sem causar diferença estatística na *deviance* residual no modelo pelo teste, a um nível de significância de 0,05. A definição dos grupos de variedades, para cada conjunto de dados, pode ser descrita pelas seguintes etapas:

- Ajustou-se o modelo final selecionado completo (Mc), ou seja, com preditor linear dado pela Equação (3.1);
- Ordenou-se, em forma decrescente de incidência predita de doença, a variável **variedade** $_{ij} = \{\mathbf{v}_{1i}, \mathbf{v}_{2i}, \mathbf{v}_{3i}, \dots, \mathbf{v}_{ni}\}$, sendo i o i -ésimo bloco e n o número de variedades no conjunto de dados. Dessa forma, agruparam-se as variedades da mais suscetível para a mais resistente;
- Criou-se uma variável **grupo** $_{ik} = \{\mathbf{g}_{1i}, \mathbf{v}_{3i}, \dots, \mathbf{v}_{ni}\}$ com $(n - 1)$ níveis, em que $\mathbf{g}_{1i} = \{\mathbf{v}_{1i}, \mathbf{v}_{2i}\}$ representava o primeiro grupo de variedades;
- Ajustou-se o modelo reduzido (Mr) com mesma distribuição e função de ligação, mas preditor linear dado por:

$$\eta_{ij} = \log \left(\frac{\pi_{ik}}{1 - \pi_{ik}} \right) = \alpha + \gamma_i + \delta_k$$

em que $i = 1, 2, \dots, 5$ e $k = 1, 2, \dots, (n - 1)$, α era a constante, γ_i era o efeito do i -ésimo bloco e δ_k o efeito do k -ésimo grupo.

- Realizou-se um teste de razão de verossimilhanças para os modelos encaixados $Mc \subset Mr$, como apresentado na Seção 2.1.3;
- Aplicou-se a seguinte regra de decisão:
 - Para p -valor $< 0,05$, rejeita-se a hipótese de que o efeito do parâmetro adicional é zero e conclui-se que as variedades v_1 e v_2 não podem ser agrupadas a um nível de significância $\alpha = 0,05$.
 - Para p -valor $\geq 0,05$, não se rejeita a hipótese nula e conclui-se que as variedades podem ser agrupadas, sem causar um aumento significativo na *deviance* residual.

- g) Fez-se uma nova definição para a variável **grupo**_{ik}, seguindo a regra aplicada em f):
- i. Para $p\text{-valor} < 0,05$, **grupo**_{ik} = {**g**_{1i}, **g**_{2i}, **v**_{4i}, ..., **v**_{ni}} com $(n - 1)$ níveis, em que **g**_{1i} = {**v**_{1i}} representa o primeiro grupo de variedades e **g**_{2i} = {**v**_{2i}, **v**_{3i}}, o segundo;
 - ii. Para $p\text{-valor} \geq 0,05$, **grupo**_{ik} = {**g**_{1i}, **v**_{4i}, ..., **v**_{ni}} com $(n-2)$ níveis, em que **g**_{1i} = {**v**_{1i}, **v**_{2i}, **v**_{3i}} representa o primeiro grupo de variedades;
- h) Repetiram-se as etapas d) a g), considerando as novas definições da variável **grupo**_{ik} modificada a cada iteração até a variedade **v**_n ser testada dentro de algum agrupamento.

Como os grupos foram definidos, atribuindo-se sequencialmente variedades por ordem decrescente em relação à incidência média de doença, realizou-se um ajuste dos grupos formados, com o intuito de se obter um modelo com a menor *deviance* residual, mantendo-se o mesmo número de grupos formados originalmente sem forçar as variedades em uma classe mais alta. Dessa forma, rebaixou-se o grupo das variedades, uma a uma, e ajustou-se, para cada iteração, o modelo reduzido novamente para obtenção das novas estatísticas. Os grupos que resultaram em um modelo com menor AIC foram finalmente estabelecidos. O ajuste dos grupos de variedades para cada conjunto de dados, pode ser descrito pelas seguintes etapas:

- a) Ordenou-se, em forma crescente de incidência predita de doença, a variável **variedade**_{ij} = {**v**_{1i}, **v**_{2i}, **v**_{3i}, ..., **v**_{ni}}, sendo *i* o *i*-ésimo bloco e *n* o número de variedades no conjunto de dados.
- b) Considerando que, nessa ordem, tinha-se a variável **grupo**_{ik} = {**g**_{1i}, **g**_{2i}, **g**_{3i}, ..., **g**_{ki}}, sendo *k* o *k*-ésimo grupo ordenado do mais resistente ao mais suscetível. Tomou-se a primeira variedade do grupo **g**₂ e moveu-a para o grupo **g**₁;
- c) Ajustou-se o mesmo modelo reduzido (*Mr*), considerando a nova variável **grupo**_{ik} e comparou-se o AIC desse modelo com o modelo anterior e utilizou-se a seguinte regra de decisão:
 - i. se o AIC(modelo novo) < AIC(modelo atual), o novo grupo definido será escolhido e, então, repetir-se-ão os passos b) e c) para as demais variedades;
 - ii. se o AIC(modelo novo) ≥ AIC(modelo atual), o grupo atual será mantido; e então repetir-se-ão os passos b) e c) para as demais variedades do próximo grupo.

3.2.4 Classificação dos grupos de variedades

Como o objetivo dos ensaios era selecionar e caracterizar clones quanto à suscetibilidade a carvão e mosaico, era necessário garantir que nenhum clone suscetível fosse classificado como resistente e que nenhum clone resistente fosse eliminado. Dessa forma, a atribuição das classes resistente, intermediária e suscetível, aos grupos de variedades formados pelo modelo, foi baseada nos resultados dos padrões de referência em cada um dos conjuntos de dados.

Sabe-se que uma variedade suscetível pode apresentar baixa (ou nula) incidência de plantas sintomáticas quando o triângulo da doença (Figura 1.1) não é atingido, evento esse chamado de escape. Por outro lado, uma variedade resistente apresentar altos índices de plantas sintomáticas é indicativo de que a pressão da doença foi muito alta. Por esse motivo, para cada um dos grupos de variedades formados pelos modelos, considerou-se a menor classe histórica do padrão de referência presente no grupo, como exemplificado na Tabela 3.4.

Tabela 3.4. Classificação dos grupos de variedades gerados pelos modelos encaixados, dada pela menor classe histórica dos padrões presentes no grupo

Variedade	Classe histórica do padrão	Grupo	Classe do grupo no ensaio
clone 1	-	g_1	resistente
clone 2	-		
padrão 1	resistente		
padrão 2	intermediária		
clone 3	-	g_2	intermediária
clone 4	-		
padrão 3	intermediária		
clone 5	-	g_3	intermediária
padrão 4	intermediária		
padrão 5	suscetível		
clone 6	-	g_4	suscetível
padrão 6	suscetível		

Nos conjuntos de dados em que a incidência de doença geral do ensaio era muito baixa, foi necessário avaliar a capacidade de diferenciação do ensaio, pois mesmo o modelo sendo capaz de gerar grupos de variedades, não necessariamente os grupos formados representavam, de fato, diferentes classes, uma vez que dentro das categorias resistente, intermediária e suscetíveis podem existir mais níveis de resistência, como apresentado na Tabela 1.1. Considerando isso, utilizou-se a escala apresentada por CHAO et al. (1989), descrita na Tabela 1.2, para estabelecer os limites mínimos de incidência predita para que um grupo recebesse as categorias intermediárias e suscetíveis, independentemente do resultado dos padrões, como exemplificado na Tabela 3.5.

Tabela 3.5. Classificação dos grupos de variedades, dada pela menor classe histórica dos padrões presentes no grupo, com aplicação dos limites estabelecidos pela escala de CHAO et al. (1989).

Variedade	Classe histórica do padrão	Grupo	% média predita de plantas infectadas	Classe do grupo no ensaio
clone 1	-	g_1	0%	resistente
clone 2	-			
padrão 1	resistente			
padrão 2	intermediária			
clone 3	-	g_2	10%	resistente
clone 4	-			
padrão 3	intermediária			
clone 5	-	g_3	12%	intermediária
padrão 4	intermediária			
padrão 5	suscetível			
clone 6	-	g_4	15%	intermediária
padrão 6	suscetível			

Assim, definiu-se que:

- se um grupo contiver pelo menos um padrão resistente, ele e todos os grupos abaixo dele serão classificados como resistentes;
- se um grupo não contiver nenhum padrão resistente, e contiver pelo menos um padrão intermediário ele será classificado como intermediário;
- se um grupo tiver apenas padrões suscetíveis, ele será classificado como suscetível;
- grupos com incidência predita ≤ 11 serão, automaticamente, classificados como resistentes;
- nenhum grupo com incidência predita ≤ 22 poderá ser classificado como suscetível.

Casos em que o modelo formou grupos de variedades que não continham nenhum padrão de referência, ou seja, composto apenas por clones, utilizou-se a escala apresentada por CHAO et al. (1989) para classificação.

As variedades com 0 plantas sintomáticas (desconsideradas nos modelos, por terem variância nula) foram classificadas como resistentes e variedades com 100% de plantas sintomáticas, suscetíveis. A classificação das variedades quanto à resistência a cada uma das doenças foi dada pela classe mais alta observada entre os dois estágios de corte em cada um dos ensaios. Fitotestes que apresentaram as três categorias (resistente, intermediárias e suscetíveis) foram ditos como válidos. Assim, os resultados do ensaio que não apresentou potencial de classificação (fitoteste não válido), em nenhum dos cortes, foram dados como inconclusivos e desconsideraram-se as classificações obtidas para as variedades. As classificações estabelecidas em cada um dos conjuntos de dados foram representadas graficamente, com intuito de avaliar a distribuição das variedades dentro de cada um dos grupos formados.

Posteriormente, avaliou-se a equivalência de classificação das variedades entre os dois estágios de corte para cada uma das doenças, em cada um dos ensaios.

4 RESULTADOS E DISCUSSÕES

4.1 Estrutura dos dados

As parcelas foram constituídas por 20 plantas, entretanto houve mortalidade de algumas touceiras, alterando o tamanho da unidade experimental. Plantas que morreram como consequência da doença foram contabilizadas na incidência.

No ensaio de 2016, 13% das parcelas tiveram uma redução do número de plantas, entretanto apenas 0,5% delas ficou com menos de 15 plantas. Já no ensaio de 2017, apenas 1% das parcelas sofreu uma redução, mas nenhuma delas ficou com menos de 15 plantas (Tabela 4.1). Dessa forma, pode-se ver que a frequência de parcelas com perdas de plantas é pequena, sendo razoável assumir que não causou grande impacto nos resultados gerados.

Tabela 4.1. Situação real das parcelas dos ensaios

Ensaio	Total de parcelas	Parcelas com 20 plantas	Parcelas com 15-19 plantas	Parcelas com menos que 15 plantas
2016 - Planta	520	87%	12,5%	0,5%
2016 - Soca	520	87%	12,5%	0,5%
2017 - Planta	270	100%	0%	0%
2017 - Soca	270	99%	1%	0%

Nota-se, a partir da Tabela 4.2, uma alta proporção de zeros no conjunto de dados dos dois ensaios, nos dois estágios de corte (planta e soca). Isso já era esperado pois para o aparecimento de sintomas, é imprescindível atingir os três princípios básicos da doença (ambiente favorável, hospedeiro suscetível e patógeno existente) e como os ensaios foram conduzidos em campo, onde não se tem controle das condições ambientais, é comum que algumas parcelas não apresentem sintomas (ou apresentam sintomas baixos), mesmo para variedades suscetíveis. Além disso, provavelmente, tem-se a presença de variedades resistentes.

Tabela 4.2. Porcentagem de zeros observados nos conjuntos de dados

Tratamentos	2016				2017			
	Carvão		Mosaico		Carvão		Mosaico	
	Planta	Soca	Planta	Soca	Planta	Soca	Planta	Soca
Clones	61,9	54,5	86,7	54,7	90,0	89,6	64,6	59,6
Pad. Resist.	53,3	66,7	100,0	100,0	86,7	86,7	80,0	70,0
Pad. Interm.	46,7	20,0	80,0	80,0	60,0	53,3	70,0	80,0
Pad. Suscet.	0,0	6,7	56,0	0,0	5,0	0,0	20,0	0,0
Todos	59,4	52,5	85,4	53,5	81,6	80,7	60,4	54,1

Algumas variedades possuem variância nula, ou seja, não apresentaram nenhuma planta sintomática para determinada doença ou, por outro lado, todas as plantas foram infectadas pela doença (Tabelas A.1 a A.4). Dessa forma, optou-se por não considerar tais variedades no ajuste do modelo para o respectivo conjunto de dados, uma vez que não trarão nenhuma contribuição, além de causarem problemas de convergência. Na Tabela 4.3, tem-se o número de variedades consideradas no ajuste dos modelos.

Nota-se que mesmo desconsiderando os dados das variedades com variância nula, tem-se alta proporção de zeros (Tabela 4.4 e Figuras 4.1 e 4.2). Devido ao excesso de zeros no conjunto de dados, somado à alta variabilidade (Figuras 4.4 a 4.7) têm-se indícios de problemas de superdispersão. Entretanto, essa questão também será avaliada na análise de resíduos dos modelos ajustados.

Tabela 4.3. Número de variedades consideradas nos modelos

Conjunto de dados	Total de tratamentos	Tratamentos no modelo - carvão	Tratamentos no modelo - mosaico
Plantio 2016 - planta	104	78	41
Plantio 2016 - soca	104	79	65
Plantio 2017 - planta	54	20	35
Plantio 2017 - soca	54	17	34

Tabela 4.4. Porcentagem de zeros observados nos conjuntos de dados, sem as variedades com variância nula

Tratamentos	2016				2017			
	Carvão		Mosaico		Carvão		Mosaico	
	Planta	Soca	Planta	Soca	Planta	Soca	Planta	Soca
Clones	48,3	39,2	64,0	32,2	63,3	48,9	40,0	28,8
Pad. Resist.	30,0	50,0	-	-	80,0	80,0	60,0	40,0
Pad. Interm.	46,7	20,0	60,0	60,0	40,0	30,0	70,0	80,0
Pad. Suscet.	0,0	6,7	56,0	0,0	5,0	0,0	20,0	0,0
Todos	45,9	37,5	62,9	30,2	51,0	38,8	38,9	27,1

Distribuição da incidência de carvão **Figura 4.1.** Distribuição da incidência de carvão dos ensaios e cortes, sem variedades com variância nula.

Figura 4.2. Distribuição da incidência de mosaico dos ensaios e cortes, sem variedades com variância nula.

4.2 Análise exploratória

Pode ser observado, por meio da Figura 4.3 que:

- Para ambas as doenças, a incidência de sintomas é, proporcionalmente, maior em cana soca. Isso pode ter acontecido porque as plantas ficaram mais tempo exposta à doença em campo;
- O ensaio de 2016, aparentemente, tinha mais clones suscetíveis que o ensaio de 2017. Isso pode ter acontecido pelo fato de o número de clones testados em 2016 ser quase duas vezes maior do que em 2017, ou seja, não houve um filtro prévio de variedades;
- Aparentemente, os ensaios estão mostrando diferenças entre os padrões resistentes, intermediários e suscetíveis quanto à incidência de sintomas. Entretanto, para mosaico os padrões intermediários quase não apresentaram sintomas (ambos ensaios) e para carvão; o mesmo observa-se para o ensaio de 2017.

Nas Figuras 4.4 a 4.7, percebe-se que a variação dos dados é maior para variedades com incidência de sintomas próximas a 50%. Já para variedades em que a incidência de sintomas está próxima de 100%, a variação dos dados tende a ser menor. Isso é esperado em dados desse tipo, pois como visto na Figura 2.2, a variância para dados binomiais não é constante e depende da probabilidade de sucesso e insucesso. Entretanto, pode ser observado que, para os dados de carvão no ensaio de 2016 de cana soca, há a presença de muitos zeros, além de uma alta variabilidade em variedades com altas incidências

de sintomas. A alta variabilidade parece ser da mesma forma para os dados de mosaico, para ambos os ensaios em cana soca, com exceção de que não há zeros observados para variedades com incidência próximas a 50% e acima.

Como visto na Seção 2.2.2, a variância da distribuição de probabilidade da binomial não é constante e, por esse motivo, é muito comum que variedades resistentes e suscetíveis tenham uma menor variabilidade entre as observações de campo, em relação às intermediárias, uma vez que suas médias, geralmente, encontram-se nos extremos da função de distribuição (Figura 4.8).

Aparentemente, as variedades padrões tendem a manter ou aumentar a incidência de sintomas da cana planta para soca, em ambas as doenças, com exceção da variedade REF_4 que apresentou uma redução de incidência média de carvão no ensaio de 2016 (Figuras 4.9 e 4.10). Dessa forma, essa variedade é um fonte de investigações futuras para entendimento dos motivos do evento, ou até mesmo, possível erro nos dados para tal variedade.

Observa-se que o mesmo acontece para os clones, uma vez que quase todos os pontos ficaram acima (ou muito próximos) da linha diagonal (função identidade) nos gráficos da Figura 4.11, indicando que os dados da cana soca são os mais importantes para classificação dos clones, pois há maior possibilidade de discriminação entre eles. As variedades REF_4 e VAR_2227 foram as que ficaram mais abaixo da linha diagonal indicando, mais uma vez, a importância na investigação da origem dos dados para essas variedades.

Além disso, nota-se que já com os dados da cana planta, pode-se ter uma noção do comportamento das variedades quanto à suscetibilidade a carvão e mosaico. Entretanto, a variedade não apresentar sintomas em cana planta, não necessariamente indica que ela seja resistente.

Figura 4.3. Distribuição da incidência média de mosaico e carvão para ensaios 2016 e 2017, por categoria e estágio de corte

Figura 4.4. Distribuição da incidência de carvão (ensaio de 2016). As variedades foram ordenadas pela média de incidência de carvão e as cores representam: clones(cinza), padrões resistentes(verde), padrões intermediários(amarelo) e padrões suscetíveis(vermelho). As variedades com variância nula em ambos os cortes foram desconsideradas.

Figura 4.5. Distribuição da incidência média de carvão (ensaio de 2017). As variedades foram ordenadas pela média de incidência de carvão e as cores representam: clones(cinza), padrões resistentes(verde), padrões intermediários(amarelo) e padrões suscetíveis(vermelho). As variedades com variância nula em ambos os cortes foram desconsideradas.

Figura 4.6. Distribuição da incidência média de mosaico (ensaio de 2016). As variedades foram ordenadas pela média de incidência de mosaico e as cores representam: clones(cinza), padrões resistentes(verde), padrões intermediários(amarelo) e padrões suscetíveis(vermelho). As variedades com variância nula em ambos os cortes foram desconsideradas.

Figura 4.7. Distribuição da incidência média de mosaico (ensaio de 2017). As variedades foram ordenadas pela média de incidência de mosaico e as cores representam: clones(cinza), padrões resistentes(verde), padrões intermediários(amarelo) e padrões suscetíveis(vermelho). As variedades com variância nula em ambos os cortes foram desconsideradas.

Figura 4.8. Representação da função de ligação *logit* à esquerda e função de variância, à direita

Figura 4.9. Incidência média de carvão e interação dos dois cortes dos padrões de referência.

Figura 4.10. Incidência média de mosaico e interação dos dois cortes dos padrões de referência.

Figura 4.11. Dispersão entre a incidência média de carvão e mosaico em cana planta e soca (ensaios 2016 e 2017). A linha diagonal representa a função identidade.

4.3 Modelagem

Ajustando-se os modelos binomiais, com preditor linear dado por (3.1), foram obtidos os resultados apresentados na Tabela 4.5 para carvão e Tabela 4.6 para mosaico.

Tabela 4.5. Estatísticas do modelo Binomial logístico para incidência de carvão

Conjunto de dados	Deviance	g.l.	$\hat{\phi}$	$\chi^2_{(n-p;0,05)}$	$p - valor$
Plantio 2016 - planta	412,9	308	1,34	349,9	0,00
Plantio 2016 - soca	693,2	312	2,22	354,2	0,00
Plantio 2017 - planta	80,2	76	1,06	97,4	0,35
Plantio 2017 - soca	94,6	64	1,48	83,7	0,01

Tabela 4.6. Estatísticas do modelo Binomial logístico para incidência de mosaico

Conjunto de dados	Deviance	g.l.	$\hat{\phi}$	$\chi^2_{(n-p;0,05)}$	$p - valor$
Plantio 2016 - planta	210,7	160	1,32	190,5	0,00
Plantio 2016 - soca	620,4	268	2,31	307,2	0,00
Plantio 2017 - planta	146,6	136	1,08	164,2	0,25
Plantio 2017 - soca	279,3	132	2,12	159,8	0,00

Observou-se que, apenas para os dados de cana planta do ensaio de 2017, os modelos tiveram indicativos de bom ajuste para incidência de carvão e mosaico com $p > 0,05$ e parâmetros de dispersão estimados muito próximos do teórico ($\hat{\phi} = 1,06$ e $\hat{\phi} = 1,08$). Entretanto, notou-se por meio dos gráficos semi-normais de probabilidades (Figuras A.1 e A.6), que muitos pontos ficaram fora do envelope, sendo necessária uma maior investigação.

Nota-se que, para o caso acima mencionado, a maioria dos pontos que ficaram fora do envelope estão muito próximas de zero. Como cada observação (parcela) possui no máximo 20 plantas observadas, acredita-se que quando apenas uma planta, de apenas uma repetição, apresenta sintoma, a variância para esse tratamento poderá ser inflacionada, uma vez que temos apenas 5 repetições por tratamento. Dessa forma, reuniram-se as variedades cujos dados possuem esta configuração (Tabela 4.7) e ajustou-se o modelo binomial novamente. Para o novo modelo ajustado, os parâmetros estimados passaram a ser $\hat{\phi} = 0,97$ para carvão e $\hat{\phi} = 1,03$ para mosaico e agora, o gráfico de probabilidade semi-normais confirmam um bom ajuste do modelo (Figura 4.12). Como a proporção de plantas infectadas para essas variedades é muito baixa e, provavelmente, seriam classificadas como resistentes, seria possível seguir as análises com o modelo reduzido. Entretanto, seguiu-se com ajustes de outros modelos para fins de comparação de modelos.

Já para os demais conjuntos de dados, notou-se a presença de superdispersão, uma vez que o desvio residual foi muito maior do que os números de graus de liberdade. Além disso, a superdispersão foi ainda maior para os dados de cana soca, em que as incidências de plantas infectadas são maiores, confirmando as observações vistas nas análises exploratórias.

Tabela 4.7. Lista de variedades com apenas uma planta sintomática em apenas uma das repetições

Conjunto de dados	Variedades
Ensaio de 2017 - planta (carvão)	REF_3, REF_4, REF_5, VAR_1560, VAR_1591, VAR_1779, VAR_2410, VAR_2421
Ensaio de 2017 - planta (mosaico)	REF_4, VAR_1480, VAR_1719, VAR_2044, VAR_2401, VAR_2402, VAR_2405

Figura 4.12. Gráficos semi-normais para modelos binomiais ajustados para incidência de carvão à esquerda, e mosaico à direita. Ajuste agrupando variedades da Tabela 4.7

Dessa forma, seguindo-se na investigação, ajustaram-se outros modelos cujos valores de desvio global, AIC e BIC estão apresentadas nas Tabelas 4.8 e 4.9. Tais estatísticas foram utilizadas como critério de seleção dos modelos, juntamente com a aderência dos dados aos envelopes simulados nos gráficos de distribuição de probabilidade semi-normal (Figuras A.1 a A.10).

Assim, tem-se que:

- Para os dados do ensaio de 2016 de cana planta, optou-se pelo modelo ZIBI para carvão e BB para mosaico, pois apresentaram menor desvio global, AIC e BIC dentre os modelos que se ajustaram pelo gráfico semi-normal de probabilidades. Houve um resíduo para a variedade VAR_1464 (carvão) e outro para a variedade VAR_2298 (mosaico) que mesmo estando dentro dos limites esperados (-3 a 3), parecem estar se comportando como outliers (Figuras 4.13 e 4.14). Como os sintomas não desaparecem facilmente em campo, em casos como esse, o ideal seria voltar ao campo para confirmar a veracidade de tais observações. Como isso não foi possível nesse caso, seguiu-se considerando todos os dados.
- Para os dados do ensaio de 2016 de cana soca, optou-se pelo modelo ZIBB para carvão e BB para mosaico, pois apresentaram menor desvio global, AIC e BIC dentre os modelos testados. As variedades VAR_2271 e VAR_2297 tiveram uma alta variabilidade entre as repetições para mosaico, notada por seus resíduos (Figuras 4.15 e 4.16).
- Uma vez que o objetivo do trabalho é agrupar e classificar variedades quanto a suscetibilidade à carvão e mosaico, para os dados do ensaio de 2017 de cana planta, optou-se pelo modelo binomial reduzido (unindo as variedades apresentadas em Tabela 4.7) para carvão e mosaico, já que este apresentou menores desvio, AIC e BIC e foi 100% aderente ao gráfico de probabilidades semi-normais;

- Para os dados de mosaico do ensaio de 2017 de cana soca, apenas o modelo BB teve indícios de bom ajuste por meio do gráfico semi-normal de probabilidades, além disso apresentou o menor desvio global para ambas as doenças e por isso, foi o modelo selecionado.

Tabela 4.8. Estatísticas dos modelos ajustados para incidência de carvão

Conjunto	Distribuição	Desvio global	AIC	BIC	% pontos fora do envelope
2016 - planta	BI	910,8	1074,8	1400,0	92%
2016 - planta	QBI	-	-	-	33%
2016 - planta	BB	910,8	1076,8	1406,0	1%
2016 - planta	ZIBI	908,1	1074,1	1403,3	3%
2016 - planta	ZIBB	912,4	1080,4	1413,6	7%
2016 - soca	BI	1344,3	1510,3	1840,6	100,0%
2016 - soca	QBI	-	-	-	39%
2016 - soca	BB	1266,2	1434,2	1768,4	3%
2016 - soca	ZIBI	1300,0	1468,0	1802,2	58%
2016 - soca	ZIBB	1259,1	1429,1	1767,3	0%
2017 - planta	BI	194,3	242,3	304,9	34,0%
2017 - planta	BI(reduzido)	194,3	228,3	272,6	1%
2017 - planta	QBI	-	-	-	26%
2017 - planta	BB	195,0	245,0	310,1	16%
2017 - planta	ZIBI	194,3	244,3	309,5	5%
2017 - planta	ZIBB	195,4	247,4	315,1	5%
2017 - soca	BI	220,5	262,5	313,8	49,0%
2017 - soca	QBI	-	-	-	11%
2017 - soca	BB	219,7	263,7	317,4	1%
2017 - soca	ZIBI	220,5	264,5	318,3	7%
2017 - soca	ZIBB	219,7	265,7	321,9	7%

Pôde ser observado que modelos inflacionados de zeros foram úteis para serem ajustados a dados de carvão, enquanto que para mosaico, modelos que consideram apenas a superdispersão foram mais adequados. Conversando com especialistas em Fitopatologia, entendeu-se que para mosaico a forma de inoculação do vírus, por linhas infectoras, pode gerar uma maior variabilidade devido à colonização não homogênea dos pulgões (transmissores). Entretanto, como a disseminação ocorre naturalmente, não necessariamente resulta em plantas totalmente assintomáticas (sem excesso de zeros). Por outro lado, para carvão, tem-se como sintoma a presença ou ausência do chicote com esporos do fungo, em que sua inoculação é realizada manual e individualmente em cada planta. Dessa forma, qualquer falha na inoculação aumenta a probabilidade de a planta não apresentar sintoma, podendo gerar um excesso zeros no conjunto de dados.

Para nenhum conjunto de dados foram observadas tendências de correlação entre as observações e nem de heteroscedasticidade, uma vez que os resíduos ficaram distribuídos aleatoriamente em torno de zero, como visto nos gráficos das Figuras 4.13 a 4.20.

Tabela 4.9. Estatísticas dos modelos ajustados para incidência de mosaico

Conjunto	Distribuição	Desvio global	AIC	BIC	% pontos fora do envelope
2016 - planta	BI	377,2	467,2	616,8	96%
2016 - planta	QBI	-	-	-	23%
2016 - planta	BB	375,4	467,4	620,3	0%
2016 - planta	ZIBI	377,4	469,4	622,2	9%
2016 - planta	ZIBB	378,2	472,2	628,4	3%
2016 - soca	BI	1189,7	1327,7	1588,8	100%
2016 - soca	QBI	-	-	-	26%
2016 - soca	BB	1099,6	1239,6	1504,5	0%
2016 - soca	ZIBI	1189,4	1329,4	1594,3	88%
2016 - soca	ZIBB	1099,6	1241,6	1510,3	26%
2017 - planta	BI	411,3	489,3	612,8	20%
2017 - planta	BI(reduzido)	411,3	477,3	581,8	0%
2017 - planta	QBI	-	-	-	31%
2017 - planta	BB	413,0	493,0	619,6	2%
2017 - planta	ZIBI	411,3	491,3	617,9	0%
2017 - planta	ZIBB	412,3	494,3	624,1	3%
2017 - soca	BI	632,7	708,7	827,9	99%
2017 - soca	QBI	-	-	-	16%
2017 - soca	BB	610,4	688,4	810,7	5%
2017 - soca	ZIBI	628,9	706,9	829,2	88%
2017 - soca	ZIBB	610,4	690,4	815,9	14%

Figura 4.13. Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo binomial inflacionado (ZIBI) ajustado para incidência de carvão (ensaio de 2016 - cana planta)

Figura 4.14. Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo beta-binomial (BB) ajustado para incidência de mosaico (ensaio de 2016 - cana planta)

Figura 4.15. Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo beta-binomial inflacionado (ZIBB) ajustado para incidência de carvão (ensaio de 2016 - cana soca)

Figura 4.16. Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo beta-binomial (BB) ajustado para incidência de mosaico (ensaio de 2016 - cana soca)

Figura 4.17. Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo binomial (BI) ajustado para incidência de carvão (ensaio de 2017 - cana planta)

Figura 4.18. Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo binomial (BI) ajustado para incidência de mosaico (ensaio de 2017 - cana planta)

Figura 4.19. Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo beta-binomial (BB) ajustado para incidência de carvão (ensaio de 2017 - cana soca)

Figura 4.20. Dispersão dos resíduos por variedade, ordem de observação e valores ajustados para modelo beta-binomial (BB) ajustado para incidência de mosaico (ensaio de 2017 - cana soca)

Por meio da Figura 4.21, consegue-se avaliar o comportamento da função de variância de cada um dos modelos finais ajustados. É nítido o ganho na utilização de modelos mais elaborados para dados com superdispersão, pois eles foram capazes de captar a variação extra dos dados, sendo mais adequados do que o modelo binomial.

Figura 4.21. Comparação das funções de variância estimadas pelos modelos finais

4.4 Agrupamento de variedades

A partir dos modelos selecionados, gerou-se uma sequência de comparação de modelos encaixados, como descrito na Seção 3.2.4. Na Tabela 4.12, estão apresentadas as estatísticas do modelo completo (por variedade) e reduzido (por grupos de variedades) para cada conjunto de dados. Este modelo contém o menor número de parâmetros (grupos) possível, sem ter um aumento significativo na variabilidade residual a nível $\alpha = 0,05$, pelo teste de razão de verossimilhanças para modelos encaixados e a melhor combinação de variedades escolhidos pelo menor AIC dos modelos.

A seguir, será apresentado o passo a passo da obtenção dos grupos de variedades para dados de mosaico do ensaio de 2017 de cana soca, como aplicação do método. O agrupamento de variedades para os demais conjuntos de dados foram realizados da mesma forma.

- Ajustou-se o modelo beta-binomial com preditor linear dado pela Equação 3.1, obtendo-se as estatísticas da Tabela 4.10;
- Ordenou-se, em forma decrescente de incidência predita de doença, a variável **variedade** = {REF_6, REF_1, REF_10, VAR_2445, REF_3, VAR_2393, VAR_2421, REF_7, VAR_2395, VAR_1704, VAR_2427, REF_2, VAR_2365, VAR_2368, VAR_1591, VAR_2026, VAR_2378, VAR_1560, VAR_2437, VAR_2410, VAR_1844, VAR_1810, REF_8, VAR_1719, VAR_2436, VAR_1795, VAR_2044, VAR_2405, VAR_2367, VAR_1832, VAR_2402, REF_4, REF_5, VAR_2388}
- Os agrupamentos foram formados, testando se a inclusão das variáveis, sequencialmente, ao grupo anterior causaria perda significativa de informação, avaliada pelo teste de razão de verossimilhança para modelos encaixados. O esquema de agrupamento sequencial está ilustrado na Figura 4.22;

Tabela 4.10. Estimativas dos parâmetros do modelo BB completo
para ensaio de 2017 - soca (mosaico)

Coefficientes	Estimativa	Erro padrão	Estatística t	p-valor
Preditor de μ com ligação logística				
(Intercepto)	0,73	0,3	2,41	0,02
BLOCO2	-0,11	0,2	-0,54	0,59
BLOCO3	-0,06	0,2	-0,28	0,78
BLOCO4	-0,57	0,21	-2,76	0,01
BLOCO5	-0,16	0,2	-0,82	0,41
VARIEDADE:REF_10	-0,13	0,39	-0,33	0,74
VARIEDADE:REF_2	-1,33	0,4	-3,37	0,00
VARIEDADE:REF_3	-0,6	0,38	-1,59	0,11
VARIEDADE:REF_4	-4,89	1,05	-4,63	0,00
VARIEDADE:REF_5	-4,89	1,06	-4,63	0,00
VARIEDADE:REF_6	0,35	0,41	0,87	0,39
VARIEDADE:REF_7	-0,84	0,38	-2,19	0,03
VARIEDADE:REF_8	-3,41	0,61	-5,62	0,00
VARIEDADE:VAR_1560	-2,31	0,46	-5,01	0,00
VARIEDADE:VAR_1591	-1,84	0,42	-4,35	0,00
VARIEDADE:VAR_1704	-1,18	0,39	-3,03	0,00
VARIEDADE:VAR_1719	-3,7	0,71	-5,23	0,00
VARIEDADE:VAR_1795	-3,79	0,72	-5,27	0,00
VARIEDADE:VAR_1810	-3,33	0,57	-5,82	0,00
VARIEDADE:VAR_1832	-4,68	1,03	-4,56	0,00
VARIEDADE:VAR_1844	-3,12	0,57	-5,44	0,00
VARIEDADE:VAR_2026	-1,93	0,42	-4,55	0,00
VARIEDADE:VAR_2044	-4,18	0,77	-5,41	0,00
VARIEDADE:VAR_2365	-1,8	0,41	-4,35	0,00
VARIEDADE:VAR_2367	-4,18	0,78	-5,39	0,00
VARIEDADE:VAR_2368	-1,82	0,42	-4,33	0,00
VARIEDADE:VAR_2378	-2,29	0,45	-5,03	0,00
VARIEDADE:VAR_2388	-4,89	1,06	-4,63	0,00
VARIEDADE:VAR_2393	-0,62	0,38	-1,63	0,1
VARIEDADE:VAR_2395	-1	0,39	-2,59	0,01
VARIEDADE:VAR_2402	-4,88	1,05	-4,63	0,00
VARIEDADE:VAR_2405	-4,18	0,78	-5,39	0,00
VARIEDADE:VAR_2410	-2,99	0,52	-5,77	0,00
VARIEDADE:VAR_2421	-0,75	0,39	-1,95	0,05
VARIEDADE:VAR_2427	-1,32	0,4	-3,33	0,00
VARIEDADE:VAR_2436	-3,74	0,7	-5,32	0,00
VARIEDADE:VAR_2437	-2,96	0,51	-5,78	0,00
VARIEDADE:VAR_2445	-0,48	0,38	-1,25	0,22
Preditor de σ com ligação log				
(Intercepto)	-3,26	0,27	-11,91	0,00

Iteração	Variedade	REF_6	REF_1	REF_10	VAR_2445	REF_3	VAR_2393	VAR_2421	REF_7	VAR_2395	VAR_1704	VAR_2427	REF_2	VAR_2365	VAR_2368	VAR_1591	VAR_2066	VAR_2378	VAR_1560	VAR_2427	VAR_2410	VAR_1844	VAR_1810	REF_8	VAR_1719	VAR_2456	VAR_1795	VAR_2044	VAR_2405	VAR_2367	VAR_1832	VAR_2402	REF_4	REF_5	VAR_2388	p-valor		
1	Agrupamento 1	g1	g1	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,37		
2	Agrupamento 2	g1	g1	g1	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v16	v17	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,46			
3	Agrupamento 3	g1	g1	g1	g1	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v17	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,20				
4	Agrupamento 4	g1	g1	g1	g1	g1	v6	v7	v8	v9	v10	v11	v12	v13	v14	v18	v19	v20	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,10		
5	Agrupamento 5	g1	g1	g1	g1	g1	g1	v7	v8	v9	v10	v11	v12	v13	v14	v19	v20	v21	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,08		
6	Agrupamento 6	g1	g1	g1	g1	g1	g1	v8	v9	v10	v11	v12	v13	v14	v20	v21	v22	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,05			
7	Agrupamento 7	g1	g1	g1	g1	g1	g1	g2	g2	v9	v10	v11	v12	v13	v14	v21	v22	v23	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,13		
8	Agrupamento 8	g1	g1	g1	g1	g1	g1	g2	g2	v10	v11	v12	v13	v14	v22	v23	v24	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,17			
9	Agrupamento 9	g1	g1	g1	g1	g1	g1	g2	g2	g2	v11	v12	v13	v14	v23	v24	v25	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,18			
10	Agrupamento 10	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	v12	v13	v14	v24	v25	v26	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,18		
11	Agrupamento 11	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	v13	v14	v25	v26	v27	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,19		
12	Agrupamento 12	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	v14	v26	v27	v28	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,07		
13	Agrupamento 13	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	v27	v28	v29	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,03		
14	Agrupamento 14	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	v29	v30	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,10		
15	Agrupamento 15	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	v31	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,13		
16	Agrupamento 16	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,14		
17	Agrupamento 17	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,16		
18	Agrupamento 18	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,08		
19	Agrupamento 19	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,05		
20	Agrupamento 20	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	g3	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,04		
21	Agrupamento 21	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,07
22	Agrupamento 22	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,09	
23	Agrupamento 23	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	v25	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,11	
24	Agrupamento 24	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	v26	v27	v28	v29	v30	v31	v32	v33	v34	0,14	
25	Agrupamento 25	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	v27	v28	v29	v30	v31	v32	v33	v34	0,16	
26	Agrupamento 26	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	v28	v29	v30	v31	v32	v33	v34	0,17	
27	Agrupamento 27	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	v29	v30	v31	v32	v33	v34	0,19		
28	Agrupamento 28	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	v30	v31	v32	v33	v34	0,22		
29	Agrupamento 29	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	v31	v32	v33	v34	0,21			
30	Agrupamento 30	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	v32	v33	v34	0,20			
31	Agrupamento 31	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	v33	v34	0,19	0,20	0,20		
32	Agrupamento 32	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	v34	0,20	0,20	0,20		
33	Agrupamento 33	g1	g1	g1	g1	g1	g1	g2	g2	g2	g2	g2	g2	g2	g2	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	g3	0,20		

Figura 4.22. Representação do esquema de agrupamento sequencial das variedades ordenadas pela incidência predita em forma decrescente, por teste de razão de verossimilhança para modelos encaixados.

- d) Definiu-se o **Agrupamento 33** como o grupo de variedade inicial e após tal definição, verificou-se a necessidade de ajuste nos grupos formados, com intuito de obter modelo com menor *deviance* residual. Os novos agrupamentos testados estão apresentados na Figura 4.23 e a escolha do agrupamento final foi feita a partir do modelo com menor AIC.

Variedade	VAR_2388	REF_5	REF_4	VAR_2402	VAR_1832	VAR_2405	VAR_2367	VAR_2044	VAR_1795	VAR_2436	VAR_1719	REF_8	VAR_1810	VAR_1844	VAR_2410	VAR_2437	VAR_1560	VAR_2378	VAR_2026	VAR_1591	VAR_2368	VAR_2365	REF_2	VAR_2427	VAR_1704	VAR_2395	REF_7	VAR_2421	VAR_2393	REF_3	VAR_2445	REF_10	REF_1	REF_6	AIC
Agrupamento 33 (atual)	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g3	g3	g3	g3	g3	g3	g3	g2	g2	g2	g2	g2	g2	g2	g1	g1	g1	g1	g1	g1	664,67
Agrupamento 34	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g3	g3	g3	g3	g3	g3	g2	g2	g2	g2	g2	g2	g2	g1	g1	g1	g1	g1	g1	665,49
Agrupamento 35	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g3	g3	g3	g3	g3	g3	g3	g3	g2	g2	g2	g2	g2	g2	g1	g1	g1	g1	g1	g1	661,93
Agrupamento 36	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g3	g3	g3	g3	g3	g3	g3	g3	g3	g2	g2	g2	g2	g2	g1	g1	g1	g1	g1	g1	666,77
Agrupamento 37	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g4	g3	g3	g3	g3	g3	g3	g3	g3	g2	g2	g2	g2	g2	g2	g1	g1	g1	g1	g1	g1	666,34

Figura 4.23. Representação dos ajustes nos grupos formados, para escolha daquele que tiver menor *deviance* residual.

O modelo com o menor AIC foi para o **Agrupamento 35**, cujos parâmetros estimados e suas estatísticas encontram-se na Tabela 4.11. Assim, esse foi o agrupamento das variedades estabelecido para os dados de mosaico para o ensaio de 2017, cana soca. As tabelas com parâmetros estimados e estatísticas para os modelos com agrupamentos finais dos demais conjuntos de dados encontram-se nas Tabelas A.5 a A.11.

Tabela 4.11. Estimativas dos parâmetros do modelo BB para ensaio de 2017 - soca (mosaico)

Coefficientes	Estimativa	Erro padrão	Estatística t	p-valor
Preditor de μ com ligação logística				
(Intercept)	0,48	0,18	2,61	0,01
BLOCO2	-0,09	0,21	-0,45	0,66
BLOCO3	-0,05	0,22	-0,22	0,83
BLOCO4	-0,55	0,22	-2,51	0,01
BLOCO5	-0,16	0,21	-0,76	0,45
GRUPO2	-0,81	0,17	-4,69	0,00
GRUPO3	-1,94	0,18	-10,63	0,00
GRUPO4	-3,68	0,23	-16,15	0,00
Preditor de σ com ligação log				
(Intercept)	-2,77	0,21	-13,38	0

Na Tabela 4.12, têm-se as comparações entre modelos completos (com todas as variedades) e reduzidos (agrupamento de variedades) para todos os conjuntos de dados. Nota-se que os dados de cana soca do ensaio de 2016 foi o que resultou em maior número de grupos de variedades, para ambas as doenças, o que faz sentido uma vez que o ensaio tem um número de variedades testadas maior em relação ao 2017 e que em cana soca as incidências de doença gerais do ensaio é maior, aumentando a chance de diferenciação dos materiais. Os dados de mosaico para cana planta do ensaio de 2016 gerou apenas 2 grupos de variedades. Esse número reduzido já era esperado, pois com a análise exploratória já se observou um nível muito baixo de infestação no ensaio (baixo poder de diferenciação). Além disso, todos os modelos finais estabelecidos não tiveram perda significativa de informação em relação ao modelo completo, pelo teste de razão de verossimilhança e dessa forma não há indícios de que as variedades agrupadas possuem diferença de incidência de doença entre si. Os agrupamentos finais de cada conjunto de dados, para cada uma das doenças encontram-se nas tabelas A.12 a A.16.

Tabela 4.12. Estatísticas dos modelos completo (por variedade) e reduzido (por grupo de variedades)

Doença	Conjunto	Modelo	Modelo completo			Modelo reduzido			p-value	Grupos de variedades
			AIC	$\hat{l}_n$	g.l.	AIC	$\hat{l}_p$	g.l.		
Carvão	2016 - planta	ZIBI	1074,1	-454,0	308	995,4	-489,72	382	0,60	3
	2016 - soca	ZIBB	1429,1	-629,5	312	1338,6	-658,3	384	0,92	5
	2017 - planta	BI	242,3	-97,2	76	214,5	-100,3	93	0,99	3
	2017 - soca	BB	263,7	-109,8	64	249,7	-115,8	76	0,53	4
Mosaico	2016 - planta	BB	467,4	-187,7	160	415,7	-200,9	198	0,94	2
	2016 - soca	BB	1239,6	-549,8	268	1175,2	-577,6	315	0,68	5
	2017 - planta	BI	489,3	-205,7	136	443,0	-214,5	168	0,98	3
	2017 - soca	BB	688,4	-305,2	132	661,9	-322,0	161	0,30	4

4.5 Classificação dos grupos de variedades

A presença de plantas sintomáticas no fitoteste depende dos três vértices do triângulo da doença (Figura 1.1) e a resistência da planta em relação a doença é uma delas (falha no vértice “hospedeiro”). Variedades resistentes podem apresentar baixa (ou nula) incidência de plantas sintomáticas, mesmo em

situações de alta pressão da doença e, por esse motivo, os padrões de referência são de grande importância nos fitotestes, pois a partir deles consegue-se avaliar a qualidade do ensaio em relação ao potencial de classificação de variedades resistentes, intermediárias e suscetíveis. Se em um fitoteste os padrões suscetíveis apresentarem baixa incidência de doença, significa que a pressão de doença está muito baixa e o ensaio não foi efetivo para classificação das variedades, pois não está explorando o potencial da doença nas plantas.

Em ambos os ensaios, observou-se que:

- para carvão e mosaico, em cana planta, foram definidos apenas grupos de padrões resistentes e intermediários (Tabelas 4.13, 4.15, 4.17 e 4.19). Nesse caso, não seria possível utilizar apenas os dados de cana planta para classificação das variedades. Já em cana soca, também houve grupos de padrões suscetíveis (Tabelas 4.14, 4.16, 4.18 e 4.20). Sendo assim, utilizando-se as classificações máximas de cada um dos cortes, foi possível dizer que ambos os ensaios, para ambas as doenças foram válidos;
- nota-se que, para o carvão, os padrões de referência ficaram distribuídos por toda a curva, indicando que os padrões foram representativos (Figuras 4.25 e 4.27). Já para o mosaico, não há padrões intermediários em toda extensão da curva, indicando que a lista de padrões utilizados necessita-se de ajuste, para aumentar representatividade dos ensaio para todas as classes de resistência (Figura 4.29 e 4.31).
- Tanto para carvão como mosaico, em cana planta, não houve padrões suscetíveis em toda extensão da curva, reforçando que somente com os dados de cana planta, pode-se subestimar o potencial de doença das variedades (Figuras 4.24, 4.26, 4.28, 4.30).

Tabela 4.13. Atribuição das classes de resistência a carvão para grupos do ensaio de 2016 (cana planta), em que R: resistente, I: intermediário e S: suscetível

Variedade padrão	Classe histórica	Predito da variedade	Grupo	Predito do grupo	Classe do grupo	Observação
REF_1	2.I	2,1%	g3	3%	R	presença de padrão resistente
REF_3	2.I	2,8%				
REF_5	1.R	4,2%				
REF_7	3.S	8,0%	g2	10%	R	presença de padrão resistente
REF_2	3.S	14,1%				
REF_4	1.R	14,6%				
REF_8	2.I	18,0%	g1	25%	I	presença de padrão intermediário e predito do grupo >11,0%
REF_6	3.S	32,3%

Figura 4.24. Representação dos grupos de variedades gerados por modelos ZIBI encaixados para dados de carvão para o ensaio de 2016 (cana planta). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.

Tabela 4.14. Atribuição das classes de resistência a carvão para grupos do ensaio de 2016 (cana soca), em que R: resistente, I: intermediário e S: suscetível

Variedade padrão	Classe histórica	Predito da variedade	Grupo	Predito do grupo	Classe do grupo	Observação
REF_4	1.R	1,4%	g5	2%	R	presença de padrão resistente
REF_1	2.I	5,5%	g4	9%	R	presença de padrão resistente
REF_3	2.I	8,0%				
REF_5	1.R	11,9%				
REF_7	3.S	27,0%	g3	36%	I	presença de padrão intermediário e predito do grupo >11,0%
REF_2	3.S	39,9%				
REF_8	2.I	47,1%				
Clones	-	-	g2	76%	S	predito do grupo >22,0%
REF_6	3.S	92,5%	g1	93%	S	predito do grupo >22,0%

Figura 4.25. Representação dos grupos de variedades gerados por modelos ZIBB encaixados para dados de carvão para o ensaio de 2016 (cana soca). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.

Tabela 4.15. Atribuição das classes de resistência a carvão para grupos do ensaio de 2017 (cana planta), em que R: resistente, I: intermediário e S: suscetível

Variedade padrão	Classe histórica	Predito da variedade	Grupo	Predito do grupo	Classe do grupo	Observação
REF_3	2.I	1,0%	g3	1%	R	presença de padrão resistente
REF_4	1.R	1,0%				
REF_5	1.R	1,0%				
REF_2	3.S	5,0%	g2	3%	R	predito do grupo $\leq 11\%$
REF_8	2.I	8,0%				
REF_7	3.S	15,0%	g1	15%	I	predito do grupo $>11,0\%$ e $\leq 22\%$
REF_6	3.S	18,0%				
REF_10	3.S	21,0%

Figura 4.26. Representação dos grupos de variedades gerados por modelos BI encaixados para dados de carvão para o ensaio de 2017 (cana planta). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.

Tabela 4.16. Atribuição das classes de resistência a carvão para grupos do ensaio de 2017 (cana soca), em que R: resistente, I: intermediário e S: suscetível

Variedade padrão	Classe histórica	Predito da variedade	Grupo	Predito do grupo	Classe do grupo	Observação
REF_4	1.R	1,2%	g4	2%	R	presença de padrão resistente
REF_5	1.R	1,7%				
REF_3	2.I	3,1%				
REF_2	3.S	10,7%	g3	12%	I	presença de padrão intermediário e predito do grupo >11,0%
REF_8	2.I	12,2%				
REF_7	3.S	21,7%				
REF_6	3.S	39,9%	g2	40%	S	presença de padrão suscetível e predito do grupo >22,0%
REF_10	3.S	85,3%	g1	86%	S	presença de padrão suscetível e predito do grupo >22,0%

Figura 4.27. Representação dos grupos de variedades gerados por modelos BB encaixados para dados de carvão para o ensaio de 2017 (cana soca). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.

Tabela 4.17. Atribuição das classes de resistência a mosaico para grupos do ensaio de 2016 (cana planta), em que R: resistente, I: intermediário e S: suscetível

Variedade padrão	Classe histórica	Predito da variedade	Grupo	Predito do grupo	Classe do grupo	Observação
REF_1	3.S	1,7%	g2	2%	R	predito do grupo ≤ 11,0%
REF_5	2.I	2,3%				
REF_7	3.S	2,3%				
REF_2	3.S	3,1%				
REF_3	3.S	5,1%				
REF_6	3.S	17,7%	g1	12%	I	predito do grupo >11,0% e ≤ 22%

Figura 4.28. Representação dos grupos de variedades gerados por modelos BB encaixados para dados de mosaico para o ensaio de 2016 (cana planta). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.

Tabela 4.18. Atribuição das classes de resistência a mosaico para grupos do ensaio de 2016 (cana soca), em que R: resistente, I: intermediário e S: suscetível

Variedade padrão	Classe histórica	Predito da variedade	Grupo	Predito do grupo	Classe do grupo	Observação
REF_5	2.I	3,1%	g5	3%	R	predito do grupo $\leq 11,0\%$
Clones	-		g4	14%	I	predito do grupo $>11,0\%$ e $\leq 22,0\%$
Clones	-		g3	36%	S	predito do grupo $>22,0\%$
REF_3	3.S	54,7%	g2	68%	S	presença de padrão suscetível e predito do grupo $>22,0\%$
REF_2	3.S	59,3%				
REF_7	3.S	72,1%				
REF_1	3.S	90,9%	g1	92%	S	presença de padrão suscetível e predito do grupo $>22,0\%$
REF_6	3.S	92,5%

Figura 4.29. Representação dos grupos de variedades gerados por modelos BB encaixados para dados de mosaico para o ensaio de 2016 (cana soca). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.

Tabela 4.19. Atribuição das classes de resistência a mosaico para grupos do ensaio de 2017 (cana planta), em que R: resistente, I: intermediário e S: suscetível

Variedade padrão	Classe histórica	Predito da variedade	Grupo	Predito do grupo	Classe do grupo	Observação
REF_4	2.I	1,0%	g3	1%	R	presença de padrão resistente
REF_8	1.R	2,0%				
REF_7	3.S	3,0%				
REF_5	2.I	4,0%	g2	4%	R	predito do grupo <11,0%
REF_6	3.S	6,0%				
REF_2	3.S	15,0%	g1	19%	I	predito do grupo >11,0% e ≤ 22,0%
REF_1	3.S	22,0%				
REF_10	3.S	22,0%				
REF_3	3.S	23,0%

Figura 4.30. Representação dos grupos de variedades gerados por modelos BI encaixados para dados de mosaico para o ensaio de 2017 (cana planta). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.

Tabela 4.20. Atribuição das classes de resistência a mosaico para grupos do ensaio de 2017 (cana soca), em que R: resistente, I: intermediário e S: suscetível

Variedade padrão	Classe histórica	Predito da variedade	Grupo	Predito do grupo	Classe do grupo	Observação
REF_4	2.I	1,4%	g4	4%	R	presença de padrão resistente
REF_5	2.I	1,4%				
REF_8	1.R	5,7%				
Clones	-		g3	16%	I	predito do grupo >11,0% e $\leq$ 22,0%
REF_2	3.S	31,5%	g2	35%	S	presença de padrão suscetível e predito do grupo >22,0%
REF_7	3.S	43,2%				
REF_3	3.S	49,0%	g1	58%	S	presença de padrão suscetível e predito do grupo >22,0%
REF_10	3.S	60,3%				
REF_1	3.S	63,7%				
REF_6	3.S	71,4%				

Ensaio 2017 - cana soca (mosaico)

Figura 4.31. Representação dos grupos de variedades gerados por modelos BB encaixados para dados de mosaico para o ensaio de 2017 (cana soca). a) distribuição da proporção de plantas sintomáticas das variedades, dentro de cada grupo, b) distribuição dos padrões de referência dentro dos grupos e c) distribuição da proporção de plantas sintomáticas e valor predito para cada variedade. R: resistente, I: intermediário e S: suscetível.

a) Ensaio 2016, carvão

		Soca		
		1.R	2.I	3.S
Planta	1.R	83,7%	11,5%	0,0%
	2.I	0,0%	2,9%	1,9%
	3.S	0,0%	0,0%	0,0%

b) Ensaio 2017, carvão

		Soca		
		1.R	2.I	3.S
Planta	1.R	85,2%	9,3%	0,0%
	2.I	0,0%	1,9%	3,7%
	3.S	0,0%	0,0%	0,0%

c) Ensaio 2016, mosaico

		Soca		
		1.R	2.I	3.S
Planta	1.R	60,6%	10,6%	25,0%
	2.I	0,0%	0,0%	3,8%
	3.S	0,0%	0,0%	0,0%

d) Ensaio 2017, mosaico

		Soca		
		1.R	2.I	3.S
Planta	1.R	63,0%	11,1%	11,1%
	2.I	0,0%	3,7%	11,1%
	3.S	0,0%	0,0%	0,0%

Figura 4.32. % de variedades com classificação equivalente em cana planta e cana soca, para dados de carvão e mosaico dos ensaios de 2016 e 2017.

Observa-se que a classificação de resistência se manteve a mesma, ou aumentou, de cana planta para soca, em ambos os ensaios e para ambas as doenças (Figura 4.32). Sendo assim, com base nos

resultados dos ensaios avaliados nesse trabalho, uma forma de otimizar recursos, mantendo a mesma qualidade de classificação dos ensaios, seria avaliar o número de plantas sintomáticas apenas em cana soca. Entretanto, com os dados de cana planta, já é possível identificar e eliminar materiais não resistentes, gerando uma antecipação de tomadas de decisão. Além disso, ao avaliar cana planta, é possível observar o número total de plantas nas parcelas e também contabilizar as plantas mortas, por decorrência da doença.

5 CONSIDERAÇÕES FINAIS

Para os dados de cana planta do ensaio 2017, tanto para carvão quanto para mosaico, o modelo binomial apresentou um bom ajuste quando se agruparam as variedades com incidências próximas a zero para todas as repetições. Já as proporções de plantas sintomáticas obtidas nos demais conjuntos de dados, apresentaram uma dispersão maior (superdispersão) do que aquela prevista pelo modelo binomial.

Entre os modelos estudados, o binomial inflacionado de zeros foi o que apresentou melhor ajuste para os dados de carvão em cana planta do ensaio de 2016, indicando que um excesso de zeros, dado, provavelmente, por fatores como dose aplicada de inoculação de carvão, que está sujeita a erros de medição e/ou seleção de isolados do fungo. Já para os dados de cana soca do mesmo ensaio, o melhor ajuste se deu pelo modelo beta-binomial inflacionado de zeros, indicando que, além de excesso de zeros, houve uma variação extra dos dados.

Para os dados de mosaico em cana soca do ensaio de 2016 e para os dois estágios de corte do ensaio 2017, o modelo final foi o beta-binomial, indicando uma variação extra dos dados, dada, provavelmente, por fatores como aninhamento de pulgões (transmissores) nas linhas infectoras de mosaico, gerando uma inoculação não homogênea. Esse mesmo modelo teve um bom ajuste para os dados de carvão em cana soca do ensaio 2017. Outra fonte de variação extra desses dados pode ter sido dada pelo fato de que os ensaios foram instalados em campo, estando expostos a inoculações naturais das doenças, ou seja, não houve controle da inoculação de forma que ela pôde ocorrer em diferentes pressões nas parcelas.

Os agrupamentos de variedades realizados por teste de razão de verossimilhança para modelos encaixados se apresentaram satisfatórios para classificação das variedades quanto à suscetibilidade a carvão e mosaico, desde que os padrões dos ensaios sejam representativos.

Para ambos os ensaios avaliados, a classificação de resistência se manteve a mesma, ou aumentou, de cana planta para soca. Sendo assim, baseado nos resultados dos ensaios avaliados nesse trabalho, uma forma de otimizar recursos, mantendo a mesma qualidade de classificação dos ensaios, seria avaliar o número de plantas sintomáticas apenas em cana soca. Entretanto, com os dados de cana planta, já é possível identificar e eliminar materiais não resistentes, gerando uma antecipação de tomadas de decisão. Além disso, ao avaliar cana planta, é possível observar o número total de plantas nas parcelas e, também, contabilizar as plantas mortas, por decorrência da doença.

Em trabalhos futuros, sugere-se uma revisão dos padrões de referência utilizados nos ensaios, para um aumento da possibilidade de o ensaio ser representativo para todas as classes de resistência.

REFERÊNCIAS

- Ahrens, W., Cox, D., e Budhwar, G. (1990). Use of the arcsine and square root transformations for subjectively determined percentage data. *Weed Science*, 38(4/5):452—458.
- AUDE, M. I. S. (1993). Estádios de desenvolvimento da cana-de-açúcar e suas relações com a produtividade. *Ciência Rural*, 23(2):241–248.
- BATISTA, D. T. (2015). Modelos para dados de contagem com superdispersão: uma aplicação em um experimento agrônomo. Dissertação (mestrado), Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.
- BUENO, C. R. N. C. (2010). Infecção por *sporisorium scitamineum* em cana-de-açúcar: influência de variáveis ambientais e desenvolvimento do método para diagnose precoce. Tese (doutorado), Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.
- CANAOESTE (2013). Doenças da cana. Disponível em: <<http://www.canaoeste.com.br/conteudo/doencas-da-cana>>. Acesso: 24/01/2019.
- CARDOSO, G. (2017). Doenças de plantas: entenda por que elas ocorrem. Disponível em: <<http://pergunteaoagronomo.com.br/entendendo-doencas-de-plantas/>>. Acesso: 24/01/2019.
- CHAO, C., HOY, J., e MARTIN, F. (1989). Evaluation of traits associated with resistance to sugarcane smut caused by *ustilago scitaminea*. *American Society of Sugar Cane Technologists (ASSCT)*, 9(4):6–16.
- CORDEIRO, G. M. e DEMÉTRIO, C. G. B. (2013). Modelos lineares generalizados e extensões. Disponível em: <<http://www.lce.esalq.usp.br/arquivos/aulas/2013/LCE5868/livro.pdf>>. Acesso: 24/01/2019.
- CORDEIRO, G. M. e LIMA NETO, E. A. (2006). Modelos paramétricos. *Universidade Federal Rural de Pernambuco UFRPE, Departamento de Estatística e Informática*.
- DEMÉTRIO, C. G. B. e HINDE, J. (1997). Half-normal plots and overdispersion. *GLIM Newsletter*, 27(2):19–26.
- DEMÉTRIO, C. G. B. (1994). A superdispersão em dados de proporção de brotação da cana-de-açúcar e de proporção de perfilhos com carvão. Tese (doutorado), Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.
- DUNN, P. K. e SMYTH, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, pages 1–9.
- FATORETTO, M. B., MORAL, R. A., DEMÉTRIO, C. G. B., PÁDUA, C. S., MENARIN, V., ROJAS, V. M. A., D’ALESSANDRO, C. P., e DELALIBERA Jr., I. (2018). Overdispersed fungus germination data: statistical analysis using R. *Biocontrol Science and Technology*, 28(11):1034–1053.
- HUTCHINSON, P. B. e DANIELS, J. (1971). A rating scale for sugarcane characteristics. *International Society of Sugar Cane Technologists (ISSCT)*, 69(4):128–131.
- MCCULLAGH, P. e NELDER, J. A. (1989). Generalized linear models.
- MONZÓON MONTROYA, A. G. (2009). Inferência e diagnóstico em modelos para dados de contagem com excesso de zeros. Dissertação (mestrado), Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica - IMECC, Campinas.

- MORAL, R. A. (2013). Modelagem estatística e ecológica de relações tróficas em pragas e inimigos naturais. Dissertação (mestrado), Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.
- MORAL, R. M., HINDE, J., e DEMÉTRIO, C. G. B. (2017). Half-normal plots and overdispersed models in R : The hnp package. *Journal of Statistical Software*, 81(10):1–23.
- PAULA, G. A. (2013). Modelos de regressão com apoio computacional.
- RIDOUT, M., DEMÉTRIO, C. G. B., e HINDE, J. (1998). Models for count data with many zeros. *Proceedings of the XIX International Biometrics Conference*, 19:179–190.
- RIGBY, R. A., STASINOPOULOS, M., Voudouris, V., Gillian, H., e Bastiani, F. D. (2015). *Flexible Regression and Smoothing: Using GAMLSS in R*.
- RIGBY, R. A., STASINOPOULOS, M. D., HELLER, G. Z., e De BASTIANI, F. (2019). *Distributions for modeling location, scale, and shape: Using GAMLSS in R*. CRC press.
- RURAL, C. (2018). Brasil, maior produtor de cana de açúcar do mundo seguido pela Índia. Disponível em: <<https://www.comprerural.com/brasil-maior-produtor-de-cana-de-acucar-do-mundo-seguido-pela-india>>. Acesso: 24/02/2019.
- SANGUINO, A. (2012). As principais doenças da cana-de-açúcar. In *Curso à Distância Tópicos Da Cultura de Cana-De-Açúcar*. Instituto Agrônômico. Disponível em: <http://www.infobibos.com/cursocana/alunos/aulas/Aula9/Aula_9.pdf>. Acesso: 24/01/2019.
- SANTIAGO, A. D. e ROSSETO, R. (2016). Doenças fúngicas. Disponível em: <http://www.agencia.cnptia.embrapa.br/gestor/cana-de-acucar/arvore/CONTAG01_79_22122006154841.html>. Acesso: 24/01/2019.
- SANTIAGO, A. D. e ROSSETO, R. (2017). Doenças da cana. Disponível em: <http://www.agencia.cnptia.embrapa.br/gestor/cana-de-acucar/arvore/CONTAG01_55_711200516718.html>. Acesso: 24/01/2019.
- STASINOPOULOS, D. M. e RIGBY, R. A. (2007). Generalized additive models for location scale and shape (gamlss) in R . *Journal of Statistical Software*, 23(7):1–46.
- STRINGER, J., CROFT, B., BHUIYAN, S., DEOMANO, E., MAGAREY, R., COX, M., e XU, X. (2012). A new method of statistical analysis for sugarcane disease screening trials. *Proceedings of the Australian Society of Sugar Cane Technology*, 34.
- WEDDERBURN, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika, Biometrika Trust*, 61(3):439–447.

APÊNDICES

Apêndice I - Tabelas descritivas dos dados desconsiderados nos modelos

Tabela A.1. Média e desvio padrão das variedades desconsideradas nos modelos para incidência de carvão - ensaio 2016

Id	Tratamento	Planta		Soca		Desconsiderado em
		Média	Desvio P.	Média	Desvio P.	
1	VAR_2259	0,00	0,00	0,01	0,02	planta
2	VAR_2254	0,00	0,00	0,01	0,02	planta
3	VAR_2282	0,00	0,00	0,01	0,03	planta
4	VAR_2292	0,00	0,00	0,02	0,03	planta
5	VAR_1440	0,00	0,00	0,04	0,04	planta
6	VAR_2251	0,00	0,00	0,02	0,04	planta
7	VAR_1428	0,00	0,00	0,00	0,00	planta e soca
8	VAR_2226	0,00	0,00	0,00	0,00	planta e soca
9	VAR_2256	0,00	0,00	0,00	0,00	planta e soca
10	VAR_2271	0,00	0,00	0,00	0,00	planta e soca
11	VAR_2291	0,00	0,00	0,00	0,00	planta e soca
12	VAR_2296	0,00	0,00	0,00	0,00	planta e soca
13	VAR_2305	0,00	0,00	0,00	0,00	planta e soca
14	REF_9	0,00	0,00	0,00	0,00	planta e soca
15	VAR_2248	0,00	0,00	0,00	0,00	planta e soca
16	VAR_2257	0,00	0,00	0,00	0,00	planta e soca
17	VAR_2260	0,00	0,00	0,00	0,00	planta e soca
18	VAR_2261	0,00	0,00	0,00	0,00	planta e soca
19	VAR_2284	0,00	0,00	0,00	0,00	planta e soca
20	VAR_2299	0,00	0,00	0,00	0,00	planta e soca
21	VAR_1441	0,00	0,00	0,00	0,00	planta e soca
22	VAR_2228	0,00	0,00	0,00	0,00	planta e soca
23	VAR_2231	0,00	0,00	0,00	0,00	planta e soca
24	VAR_2245	0,00	0,00	0,00	0,00	planta e soca
25	VAR_2287	0,00	0,00	0,00	0,00	planta e soca
26	VAR_2303	0,00	0,00	0,00	0,00	planta e soca
27	VAR_2272	0,01	0,02	0,00	0,00	soca
28	VAR_2311	0,01	0,02	0,00	0,00	soca
29	VAR_2273	0,01	0,02	0,00	0,00	soca
30	VAR_2295	0,01	0,03	0,00	0,00	soca
31	VAR_2250	0,02	0,03	0,00	0,00	soca

Tabela A.2. Média e desvio padrão das variedades desconsideradas nos modelos para incidência de mosaico - ensaio 2016

Id	Tratamento	Planta		Soca		Desconsiderado em
		Média	Desvio P.	Média	Desvio P.	
1	VAR_2245	0,00	0,00	0,01	0,02	planta
2	VAR_2232	0,00	0,00	0,01	0,02	planta
3	VAR_2262	0,00	0,00	0,01	0,02	planta
4	VAR_2276	0,00	0,00	0,01	0,02	planta
5	VAR_2279	0,00	0,00	0,01	0,02	planta
6	VAR_2288	0,00	0,00	0,01	0,02	planta
7	VAR_2226	0,00	0,00	0,02	0,03	planta
8	VAR_2256	0,00	0,00	0,02	0,03	planta
9	VAR_2291	0,00	0,00	0,02	0,03	planta
10	VAR_2235	0,00	0,00	0,02	0,03	planta
11	VAR_2305	0,00	0,00	0,02	0,04	planta
12	VAR_2265	0,00	0,00	0,02	0,04	planta
13	VAR_2308	0,00	0,00	0,02	0,04	planta
14	VAR_2294	0,00	0,00	0,03	0,04	planta
15	VAR_2263	0,00	0,00	0,04	0,07	planta
16	VAR_2238	0,00	0,00	0,05	0,06	planta
17	VAR_2282	0,00	0,00	0,06	0,06	planta
18	VAR_2255	0,00	0,00	0,06	0,11	planta
19	VAR_2259	0,00	0,00	0,09	0,04	planta
20	VAR_396	0,00	0,00	0,09	0,07	planta
21	VAR_1464	0,00	0,00	0,23	0,18	planta
22	VAR_1428	0,00	0,00	0,30	0,13	planta
23	VAR_2296	0,00	0,00	0,31	0,14	planta
24	VAR_2266	0,00	0,00	0,32	0,18	planta
25	VAR_2270	0,00	0,00	0,46	0,23	planta
26	VAR_2272	0,00	0,00	0,49	0,14	planta
27	VAR_2271	0,00	0,00	0,53	0,36	planta
28	VAR_2310	0,00	0,00	0,63	0,28	planta
29	VAR_2301	0,00	0,00	0,74	0,16	planta
30	VAR_2239	0,00	0,00	0,88	0,12	planta
31	VAR_1440	0,00	0,00	0,00	0,00	planta e soca
32	VAR_2250	0,00	0,00	0,00	0,00	planta e soca
33	VAR_2273	0,00	0,00	0,00	0,00	planta e soca
34	VAR_2292	0,00	0,00	0,00	0,00	planta e soca
35	VAR_2311	0,00	0,00	0,00	0,00	planta e soca
36	REF_9	0,00	0,00	0,00	0,00	planta e soca
37	VAR_2248	0,00	0,00	0,00	0,00	planta e soca
38	VAR_2257	0,00	0,00	0,00	0,00	planta e soca
39	VAR_2260	0,00	0,00	0,00	0,00	planta e soca
40	VAR_2261	0,00	0,00	0,00	0,00	planta e soca
41	VAR_2284	0,00	0,00	0,00	0,00	planta e soca

Tabela A.2 continuação da página anterior

Id	Tratamento	Planta		Soca		Desconsiderado em
		Média	Desvio P.	Média	Desvio P.	
42	VAR_2299	0,00	0,00	0,00	0,00	planta e soca
43	REF_4	0,00	0,00	0,00	0,00	planta e soca
44	REF_8	0,00	0,00	0,00	0,00	planta e soca
45	VAR_1458	0,00	0,00	0,00	0,00	planta e soca
46	VAR_1836	0,00	0,00	0,00	0,00	planta e soca
47	VAR_2233	0,00	0,00	0,00	0,00	planta e soca
48	VAR_2234	0,00	0,00	0,00	0,00	planta e soca
49	VAR_2241	0,00	0,00	0,00	0,00	planta e soca
50	VAR_2244	0,00	0,00	0,00	0,00	planta e soca
51	VAR_2247	0,00	0,00	0,00	0,00	planta e soca
52	VAR_2258	0,00	0,00	0,00	0,00	planta e soca
53	VAR_2264	0,00	0,00	0,00	0,00	planta e soca
54	VAR_2267	0,00	0,00	0,00	0,00	planta e soca
55	VAR_2269	0,00	0,00	0,00	0,00	planta e soca
56	VAR_2275	0,00	0,00	0,00	0,00	planta e soca
57	VAR_2277	0,00	0,00	0,00	0,00	planta e soca
58	VAR_2281	0,00	0,00	0,00	0,00	planta e soca
59	VAR_2283	0,00	0,00	0,00	0,00	planta e soca
60	VAR_2285	0,00	0,00	0,00	0,00	planta e soca
61	VAR_2293	0,00	0,00	0,00	0,00	planta e soca
62	VAR_2302	0,00	0,00	0,00	0,00	planta e soca
63	VAR_2304	0,00	0,00	0,00	0,00	planta e soca
64	VAR_2229	0,01	0,02	0,00	0,00	soca
65	VAR_2295	0,01	0,02	0,00	0,00	soca
66	VAR_2251	0,03	0,07	0,00	0,00	soca
67	VAR_2240	0,04	0,04	1,00	0,00	soca
68	VAR_2246	0,02	0,04	1,00	0,00	soca
69	VAR_2243	0,16	0,16	1,00	0,00	soca

Tabela A.3. Média e desvio padrão das variedades desconsideradas nos modelos para incidência de carvão - ensaio 2017

Id	Tratamento	Planta		Soca		Desconsiderado em
		Média	Desvio P.	Média	Desvio P.	
1	VAR_2390	0,00	0,00	0,01	0,02	planta
2	REF_1	0,00	0,00	0,00	0,00	planta e soca
3	REF_9	0,00	0,00	0,00	0,00	planta e soca
4	VAR_1613	0,00	0,00	0,00	0,00	planta e soca
5	VAR_1643	0,00	0,00	0,00	0,00	planta e soca
6	VAR_1704	0,00	0,00	0,00	0,00	planta e soca
7	VAR_1719	0,00	0,00	0,00	0,00	planta e soca
8	VAR_1795	0,00	0,00	0,00	0,00	planta e soca
9	VAR_1810	0,00	0,00	0,00	0,00	planta e soca

Tabela A.3 continuação da página anterior

Id	Tratamento	Planta		Soca		Desconsiderado em
		Média	Desvio P.	Média	Desvio P.	
10	VAR_2026	0,00	0,00	0,00	0,00	planta e soca
11	VAR_2044	0,00	0,00	0,00	0,00	planta e soca
12	VAR_2117	0,00	0,00	0,00	0,00	planta e soca
13	VAR_2365	0,00	0,00	0,00	0,00	planta e soca
14	VAR_2368	0,00	0,00	0,00	0,00	planta e soca
15	VAR_2371	0,00	0,00	0,00	0,00	planta e soca
16	VAR_2378	0,00	0,00	0,00	0,00	planta e soca
17	VAR_2382	0,00	0,00	0,00	0,00	planta e soca
18	VAR_2388	0,00	0,00	0,00	0,00	planta e soca
19	VAR_2395	0,00	0,00	0,00	0,00	planta e soca
20	VAR_2396	0,00	0,00	0,00	0,00	planta e soca
21	VAR_2399	0,00	0,00	0,00	0,00	planta e soca
22	VAR_2401	0,00	0,00	0,00	0,00	planta e soca
23	VAR_2402	0,00	0,00	0,00	0,00	planta e soca
24	VAR_2403	0,00	0,00	0,00	0,00	planta e soca
25	VAR_2405	0,00	0,00	0,00	0,00	planta e soca
26	VAR_2406	0,00	0,00	0,00	0,00	planta e soca
27	VAR_2413	0,00	0,00	0,00	0,00	planta e soca
28	VAR_2414	0,00	0,00	0,00	0,00	planta e soca
29	VAR_2425	0,00	0,00	0,00	0,00	planta e soca
30	VAR_2426	0,00	0,00	0,00	0,00	planta e soca
31	VAR_2431	0,00	0,00	0,00	0,00	planta e soca
32	VAR_2436	0,00	0,00	0,00	0,00	planta e soca
33	VAR_2437	0,00	0,00	0,00	0,00	planta e soca
34	VAR_2445	0,00	0,00	0,00	0,00	planta e soca
35	VAR_1560	0,01	0,02	0,00	0,00	soca
36	VAR_1591	0,01	0,02	0,00	0,00	soca
37	VAR_2410	0,01	0,02	0,00	0,00	soca
38	VAR_2421	0,01	0,02	0,00	0,00	soca

Tabela A.4. Média e desvio padrão das variedades desconsideradas nos modelos para incidência de mosaico - ensaio 2017

Id	Tratamento	Planta		Soca		Desconsiderado em
		Média	Desvio P.	Média	Desvio P.	
1	VAR_2388	0,00	0,00	0,01	0,02	planta
2	VAR_1832	0,00	0,00	0,02	0,04	planta
3	REF_9	0,00	0,00	0,00	0,00	planta e soca
4	VAR_1613	0,00	0,00	0,00	0,00	planta e soca
5	VAR_1643	0,00	0,00	0,00	0,00	planta e soca
6	VAR_1779	0,00	0,00	0,00	0,00	planta e soca
7	VAR_2117	0,00	0,00	0,00	0,00	planta e soca
8	VAR_2371	0,00	0,00	0,00	0,00	planta e soca

Tabela A.4 continuação da página anterior

Id	Tratamento	Planta		Soca		Desconsiderado em
		Média	Desvio P.	Média	Desvio P.	
9	VAR_2382	0,00	0,00	0,00	0,00	planta e soca
10	VAR_2390	0,00	0,00	0,00	0,00	planta e soca
11	VAR_2396	0,00	0,00	0,00	0,00	planta e soca
12	VAR_2399	0,00	0,00	0,00	0,00	planta e soca
13	VAR_2403	0,00	0,00	0,00	0,00	planta e soca
14	VAR_2404	0,00	0,00	0,00	0,00	planta e soca
15	VAR_2406	0,00	0,00	0,00	0,00	planta e soca
16	VAR_2413	0,00	0,00	0,00	0,00	planta e soca
17	VAR_2414	0,00	0,00	0,00	0,00	planta e soca
18	VAR_2425	0,00	0,00	0,00	0,00	planta e soca
19	VAR_2426	0,00	0,00	0,00	0,00	planta e soca
20	VAR_1480	0,01	0,02	0,00	0,00	soca
21	VAR_2401	0,01	0,02	0,00	0,00	soca
22	VAR_2431	0,02	0,03	0,00	0,00	soca

Apêndice II - Gráficos semi-normal para análises individuais

Modelo Binomial logístico - carvão

Figura A.1. Gráficos semi-normais para modelos Binomiais logísticos ajustados para carvão

Modelo Quase-Binomial logístico - carvão

Figura A.2. Gráficos semi-normais para modelos Quase-binomiais ajustados para carvão

Modelo Beta-Binomial - carvão

Figura A.3. Gráficos semi-normais para modelos Beta-binomiais ajustados para carvão

Figura A.4. Gráficos semi-normais para modelos Binomiais inflacionados ajustados para carvão

Figura A.5. Gráficos semi-normais para modelos Beta-binomiais inflacionados ajustados para carvão

Figura A.6. Gráficos semi-normais para modelos Binomiais logísticos ajustados para mosaico

Figura A.7. Gráficos semi-normais para modelos Quase-binomiais ajustados para mosaico

Figura A.8. Gráficos semi-normais para modelos Beta-binomiais ajustados para mosaico

Figura A.9. Gráficos semi-normais para modelos Binomiais inflacionados ajustados para mosaico

Figura A.10. Gráficos semi-normais para modelos Beta-binomiais inflacionados ajustados para mosaico

Apêndice III - Tabelas com parâmetros estimados dos modelos reduzidos

Tabela A.5. Estimativas dos parâmetros do modelo ZIBI para ensaio 2016 - planta (carvão)

Coefficientes	Estimativa	Erro padrão	Estatística t	p-valor
Preditor de μ com ligação logística				
(Intercepto)	-1,14	0,16	-7,25	0,00
BLOCO2	-0,14	0,19	-0,75	0,46
BLOCO3	0,34	0,15	2,20	0,03
BLOCO4	0,10	0,16	0,61	0,54
BLOCO5	0,02	0,16	0,12	0,9
GRUPO2	-1,17	0,11	-10,40	0,00
GRUPO3	-2,59	0,14	-17,87	0,00
Preditor de σ com ligação logística				
(Intercepto)	-2,22	0,36	-6,14	0,00

Tabela A.6. Estimativas dos parâmetros do modelo ZIBB para ensaio 2016 - soca (carvão)

Coefficientes	Estimativa	Erro padrão	Estatística t	p-valor
Preditor de μ com ligação logística				
(Intercepto)	2,32	0,4	5,77	0,00
BLOCO2	0,26	0,18	1,45	0,15
BLOCO3	0,44	0,17	2,58	0,01
BLOCO4	0,40	0,19	2,11	0,04
BLOCO5	0,48	0,18	2,65	0,01

Table A.6 continued from previous page

Coeficientes	Estimativa	Erro padrão	Estatística t	p-valor
GRUPO2	-1,47	0,43	-3,43	0,00
GRUPO3	-3,26	0,39	-8,35	0,00
GRUPO4	-4,82	0,39	-12,23	0,00
GRUPO5	-6,27	0,42	-15,05	0,00
Preditor de σ com ligação log				
(Intercepto)	-2,76	0,13	-20,82	0
Preditor de τ com ligação logit				
(Intercepto)	-3,58	0,7	-5,12	0

Tabela A.7. Estimativas dos parâmetros do modelo Binomial para ensaio 2017 - planta (carvão)

Coeficientes	Estimativa	Erro padrão	Estatística t	p-valor
Preditor de μ com ligação logística				
(Intercept)	-1,62	0,27	-6,01	0,00
BLOCO2	0,12	0,34	0,34	0,73
BLOCO3	0,12	0,34	0,34	0,73
BLOCO4	0,06	0,35	0,17	0,86
BLOCO5	0,23	0,34	0,67	0,5
GRUPO2	-1,27	0,24	-5,23	0,00
GRUPO3	-2,79	0,29	-9,51	0,00

Tabela A.8. Estimativas dos parâmetros do modelo BB para ensaio 2017 - soca (carvão)

Coeficientes	Estimativa	Erro padrão	Estatística t	p-valor
Preditor de μ com ligação logística				
(Intercept)	1,39	0,39	3,57	0,00
BLOCO2	0,53	0,35	1,53	0,13
BLOCO3	0,75	0,35	2,16	0,03
BLOCO4	0,45	0,36	1,27	0,21
BLOCO5	0,07	0,36	0,19	0,85
GRUPO2	-2,17	0,44	-4,96	0,00
GRUPO3	-3,76	0,39	-9,55	0,00
GRUPO4	-5,70	0,47	-12,23	0,00
Preditor de σ com ligação log				
(Intercept)	-3,79	0,61	-6,22	0

Tabela A.9. Estimativas dos parâmetros do modelo BB para ensaio 2016 - planta (mosaico)

Coeficientes	Estimativa	Erro padrão	Estatística t	p-valor
Preditor de μ com ligação logística				
(Intercept)	-1,61	0,25	-6,41	0,00
BLOCO2	-0,95	0,37	-2,58	0,01

Table A.9 continued from previous page

Coeficientes	Estimativa	Erro padrão	Estatística t	p-valor
BLOCO3	-0,47	0,32	-1,46	0,15
BLOCO4	-0,31	0,31	-0,99	0,32
BLOCO5	-0,41	0,31	-1,31	0,19
GRUPO2	-1,78	0,23	-7,72	0,00
Preditor de σ com ligação log				
(Intercepto)	-3,72	0,31	-11,84	0,00

Tabela A.10. Estimativas dos parâmetros do modelo BB para
ensaio 2016 - soca (mosaico)

Coeficientes	Estimativa	Erro padrão	Estatística t	p-valor
Preditor de μ com ligação logística				
(Intercept)	2,62	0,24	10,93	0,00
BLOCO2	-0,45	0,19	-2,36	0,02
BLOCO3	-0,21	0,19	-1,10	0,27
BLOCO4	-0,29	0,19	-1,57	0,12
BLOCO5	-0,17	0,19	-0,87	0,39
GRUPO2	-1,65	0,23	-7,12	0,00
GRUPO3	-3,00	0,24	-12,25	0,00
GRUPO4	-4,09	0,25	-16,65	0,00
GRUPO5	-5,84	0,26	-22,61	0,00
Preditor de σ com ligação log				
(Intercept)	-2,28	0,13	-17,53	0,00

Tabela A.11. Estimativas dos parâmetros do modelo Binomial
para ensaio 2017 - planta (mosaico)

Coeficientes	Estimativa	Erro padrão	Estatística t	p-valor
Preditor de μ com ligação logística				
(Intercept)	-1,31	0,16	-8,34	0,00
BLOCO2	0,00	0,21	0,00	1,00
BLOCO3	0,12	0,20	0,6	0,55
BLOCO4	-0,44	0,22	-1,98	0,05
BLOCO5	-0,18	0,21	-0,85	0,4
GRUPO2	-1,32	0,15	-8,86	0,00
GRUPO3	-2,65	0,22	-12,21	0,00

Apêndice IV - Agrupamentos das variedades

Tabela A.12. Agrupamento final para conjunto de dados de car-
vão para ensaio 2016 (cana planta)

Variedade	Predito	Grupo	Variedade	Predito	Grupo	Variedade	Predito	Grupo
REF_6	32,3%	g1	VAR_2264	6,0%	g2	VAR_2300	2,1%	g3

Table A.12 continued from previous page

Variedade	Predito	Grupo	Variedade	Predito	Grupo	Variedade	Predito	Grupo
VAR_2253	32,0%	g1	VAR_2289	5,5%	g2	VAR_2240	2,1%	g3
VAR_2310	24,0%	g1	VAR_2263	5,4%	g2	REF_1	2,1%	g3
REF_8	18,0%	g1	VAR_2233	5,4%	g2	VAR_2250	2,1%	g3
VAR_1464	17,3%	g1	VAR_2275	5,4%	g2	VAR_1836	2,1%	g3
VAR_2227	16,1%	g2	VAR_2280	5,3%	g2	VAR_2236	2,1%	g3
VAR_2268	15,6%	g2	VAR_2235	5,1%	g3	VAR_2249	2,1%	g3
VAR_2274	15,0%	g2	VAR_2306	5,1%	g3	VAR_2307	2,1%	g3
REF_4	14,6%	g2	VAR_2241	4,3%	g3	VAR_2265	2,1%	g3
REF_2	14,1%	g2	VAR_2302	4,3%	g3	VAR_2285	1,2%	g3
VAR_2229	13,9%	g2	VAR_2270	4,2%	g3	VAR_2295	1,1%	g3
VAR_2255	13,2%	g2	VAR_2283	4,2%	g3	VAR_2234	1,1%	g3
VAR_1445	11,9%	g2	REF_5	4,2%	g3	VAR_2273	1,1%	g3
VAR_2238	10,7%	g2	VAR_2286	4,1%	g3	VAR_396	1,0%	g3
VAR_2288	8,9%	g2	VAR_2312	4,1%	g3	VAR_2232	1,0%	g3
VAR_2242	8,9%	g2	VAR_2246	4,1%	g3	VAR_2297	1,0%	g3
REF_7	8,0%	g2	VAR_2276	3,2%	g3	VAR_2311	1,0%	g3
VAR_2267	8,0%	g2	VAR_2247	3,2%	g3	VAR_2258	1,0%	g3
VAR_2266	8,0%	g2	VAR_2237	3,2%	g3	VAR_1458	1,0%	g3
VAR_2278	7,3%	g2	VAR_2244	3,2%	g3	VAR_2308	1,0%	g3
VAR_2279	7,3%	g2	VAR_2243	3,1%	g3	VAR_2239	1,0%	g3
VAR_2290	6,8%	g2	VAR_2309	3,1%	g3	VAR_2252	1,0%	g3
VAR_2230	6,7%	g2	REF_3	2,8%	g3	VAR_2301	1,0%	g3
VAR_2293	6,4%	g2	VAR_2281	2,6%	g3	VAR_2269	1,0%	g3
VAR_2262	6,2%	g2	VAR_2298	2,4%	g3	VAR_2272	1,0%	g3
VAR_2294	6,2%	g2	VAR_2277	2,2%	g3	VAR_2304	1,0%	g3

Tabela A.13. Agrupamento final para conjunto de dados de carvão para ensaio 2016 (cana soca)

Variedade	Predito	Grupo	Variedade	Predito	Grupo	Variedade	Predito	Grupo
REF_6	92,5%	g1	REF_5	11,9%	g4	VAR_2297	4,0%	g5
VAR_1464	75,4%	g2	VAR_2262	11,9%	g4	VAR_2307	3,1%	g5
VAR_2241	51,2%	g3	VAR_2306	11,1%	g4	VAR_2277	2,9%	g5
REF_8	47,1%	g3	VAR_2229	10,8%	g4	VAR_2265	2,9%	g5
VAR_1445	46,6%	g3	VAR_2278	10,3%	g4	VAR_2292	2,8%	g5
VAR_2310	43,0%	g3	VAR_2227	10,2%	g4	VAR_2230	2,8%	g5
VAR_2253	41,2%	g3	VAR_2242	10,0%	g4	VAR_2301	2,8%	g5
REF_2	39,9%	g3	VAR_2281	8,7%	g4	VAR_2308	2,8%	g5
VAR_2279	37,1%	g3	REF_3	8,0%	g4	VAR_2285	2,5%	g5
VAR_2268	36,4%	g3	VAR_2267	8,0%	g4	VAR_2298	2,1%	g5
VAR_2252	32,5%	g3	VAR_396	8,0%	g4	VAR_2234	1,9%	g5
VAR_2290	32,4%	g3	VAR_1836	7,4%	g4	VAR_2237	1,8%	g5
VAR_2238	31,6%	g3	VAR_2249	7,1%	g4	VAR_2269	1,8%	g5
VAR_2255	30,1%	g3	VAR_2289	7,0%	g4	VAR_2302	1,8%	g5
REF_7	27,0%	g3	VAR_2309	6,5%	g4	VAR_2240	1,8%	g5

Table A.13 continued from previous page

Variedade	Predito	Grupo	Variedade	Predito	Grupo	Variedade	Predito	Grupo
VAR_2293	24,9%	g3	VAR_2244	6,5%	g4	VAR_2251	1,8%	g5
VAR_2235	22,9%	g3	VAR_2233	6,5%	g4	VAR_2282	1,6%	g5
VAR_2288	19,5%	g4	VAR_2246	6,3%	g4	VAR_2300	1,4%	g5
VAR_2312	18,6%	g4	VAR_2258	6,1%	g4	VAR_2232	1,4%	g5
VAR_2274	18,2%	g4	VAR_2276	5,9%	g4	VAR_2304	1,4%	g5
VAR_2280	16,1%	g4	VAR_2270	5,8%	g4	REF_4	1,4%	g5
VAR_2294	15,3%	g4	REF_1	5,5%	g4	VAR_2254	1,4%	g5
VAR_2266	14,1%	g4	VAR_2239	5,5%	g4	VAR_1458	1,4%	g5
VAR_2264	13,6%	g4	VAR_1440	4,9%	g5	VAR_2243	1,4%	g5
VAR_2283	12,5%	g4	VAR_2247	4,9%	g5	VAR_2259	1,4%	g5
VAR_2263	12,5%	g4	VAR_2236	4,9%	g5			
VAR_2275	11,9%	g4	VAR_2286	4,2%	g5			

Tabela A.14. Agrupamento final para conjunto de dados de mosaico para ensaio 2016 (cana planta)

Variedade	Predito	Grupo	Variedade	Predito	Grupo
REF_6	17,7%	g1	VAR_2307	2,3%	g2
VAR_2243	15,3%	g1	VAR_2268	2,1%	g2
VAR_2289	7,7%	g1	REF_1	1,7%	g2
VAR_2306	7,0%	g1	VAR_2246	1,7%	g2
VAR_2298	5,9%	g2	VAR_2312	1,7%	g2
VAR_2300	5,7%	g2	VAR_2295	1,2%	g2
REF_3	5,1%	g2	VAR_1441	1,1%	g2
VAR_2227	4,8%	g2	VAR_1445	1,1%	g2
VAR_2240	4,3%	g2	VAR_2229	1,1%	g2
VAR_2290	3,6%	g2	VAR_2230	1,1%	g2
VAR_2274	3,4%	g2	VAR_2236	1,1%	g2
VAR_2280	3,4%	g2	VAR_2237	1,1%	g2
REF_2	3,1%	g2	VAR_2249	1,1%	g2
VAR_2228	2,3%	g2	VAR_2252	1,1%	g2
REF_7	2,3%	g2	VAR_2253	1,1%	g2
VAR_2251	2,3%	g2	VAR_2254	1,1%	g2
REF_5	2,3%	g2	VAR_2278	1,1%	g2
VAR_2231	2,3%	g2	VAR_2286	1,1%	g2
VAR_2242	2,3%	g2	VAR_2303	1,1%	g2
VAR_2287	2,3%	g2	VAR_2309	1,1%	g2
VAR_2297	2,3%	g2			

Tabela A.15. Agrupamento final para conjunto de dados de mosaico para ensaio 2016 (cana soca)

Variedade	Predito	Grupo	Variedade	Predito	Grupo	Variedade	Predito	Grupo
VAR_2289	98,1%	g1	VAR_2266	32,3%	g3	VAR_2294	3,7%	g5
VAR_2300	98,1%	g1	VAR_2296	31,5%	g3	REF_5	3,1%	g5

Table A.15 continued from previous page

Variedade	Predito	Grupo	Variedade	Predito	Grupo	Variedade	Predito	Grupo
REF_6	92,5%	g1	VAR_1428	30,9%	g3	VAR_2226	3,1%	g5
REF_1	90,9%	g1	VAR_2280	30,2%	g3	VAR_2235	3,1%	g5
VAR_2227	89,8%	g1	VAR_2274	27,9%	g3	VAR_2256	3,1%	g5
VAR_2239	88,7%	g1	VAR_1464	22,9%	g4	VAR_2286	3,1%	g5
VAR_2231	87,4%	g1	VAR_1445	21,4%	g4	VAR_2291	3,1%	g5
VAR_2306	82,0%	g1	VAR_2254	20,2%	g4	VAR_2265	1,9%	g5
VAR_2290	80,7%	g2	VAR_2252	19,7%	g4	VAR_2305	1,9%	g5
VAR_2230	75,9%	g2	VAR_2312	16,8%	g4	VAR_2308	1,9%	g5
VAR_2298	75,2%	g2	VAR_2268	15,7%	g4	VAR_2276	1,6%	g5
VAR_2301	74,8%	g2	VAR_2249	13,5%	g4	VAR_1441	1,6%	g5
REF_7	72,1%	g2	VAR_2287	11,6%	g4	VAR_2232	1,6%	g5
VAR_2297	68,9%	g2	VAR_2259	11,3%	g4	VAR_2236	1,6%	g5
VAR_2278	67,1%	g2	VAR_396	10,3%	g4	VAR_2245	1,6%	g5
VAR_2310	64,9%	g2	VAR_2228	8,7%	g4	VAR_2253	1,6%	g5
VAR_2307	63,2%	g2	VAR_2282	7,0%	g5	VAR_2262	1,6%	g5
REF_2	59,3%	g2	VAR_2238	6,0%	g5	VAR_2279	1,6%	g5
REF_3	54,7%	g2	VAR_2242	5,9%	g5	VAR_2288	1,6%	g5
VAR_2271	54,5%	g2	VAR_2237	5,4%	g5	VAR_2303	1,6%	g5
VAR_2272	49,0%	g3	VAR_2255	4,7%	g5	VAR_2309	1,6%	g5
VAR_2270	46,0%	g3	VAR_2263	4,1%	g5			

Tabela A.16. Agrupamento final para conjunto de dados de cana para ensaio 2017 (cana planta)

Variedade	Predito	Grupo
REF_10	21,0%	g1
REF_6	18,0%	g1
REF_7	15,0%	g1
REF_8	8,0%	g2
VAR_1480	8,0%	g2
REF_2	5,0%	g2
VAR_2367	4,0%	g2
VAR_2404	4,0%	g2
VAR_1832	2,0%	g3
VAR_1844	2,0%	g3
VAR_2393	2,0%	g3
VAR_2427	2,0%	g3
REF_3	1,0%	g3
VAR_1779	1,0%	g3
VAR_2410	1,0%	g3
REF_4	1,0%	g3
REF_5	1,0%	g3
VAR_1560	1,0%	g3
VAR_2421	1,0%	g3

Table A.16 continued from previous page

Variedade	Predito	Grupo
VAR_1591	1,0%	g3

Tabela A.17. Agrupamento final para conjunto de dados de carvão para ensaio 2017 (cana soca)

Variedade	Predito	Grupo
REF_10	85,3%	g1
REF_6	39,9%	g2
REF_7	21,7%	g3
REF_8	12,2%	g3
VAR_1480	11,1%	g3
REF_2	10,7%	g3
VAR_2404	10,5%	g3
VAR_1779	7,3%	g3
VAR_2367	4,3%	g4
REF_3	3,1%	g4
VAR_2427	2,3%	g4
VAR_1832	1,7%	g4
REF_5	1,7%	g4
VAR_2393	1,2%	g4
VAR_2390	1,2%	g4
VAR_1844	1,2%	g4
REF_4	1,2%	g4

Tabela A.18. Agrupamento final para conjunto de dados de mosaico para ensaio 2017 (cana planta)

Variedade	Predito	Grupo	Variedade	Predito	Grupo
REF_3	23,0%	g1	REF_7	3,0%	g3
REF_10	22,0%	g1	VAR_2410	3,0%	g3
VAR_2026	22,0%	g1	VAR_2368	3,0%	g3
REF_1	22,0%	g1	VAR_2367	2,0%	g3
VAR_2421	20,0%	g1	VAR_2436	2,0%	g3
VAR_2393	20,0%	g1	REF_8	2,0%	g3
VAR_1591	15,0%	g1	VAR_1844	2,0%	g3
REF_2	15,0%	g1	VAR_2431	2,0%	g3
VAR_2427	11,0%	g2	VAR_2402	1,0%	g3
VAR_2445	8,0%	g2	VAR_2044	1,0%	g3
VAR_1810	8,0%	g2	VAR_2401	1,0%	g3
VAR_2437	7,0%	g2	REF_4	1,0%	g3
VAR_2395	7,0%	g2	VAR_1480	1,0%	g3
VAR_1704	6,0%	g2	VAR_2405	1,0%	g3
REF_6	6,0%	g2	VAR_1719	1,0%	g3
VAR_1560	5,0%	g2			
VAR_1795	5,0%	g2			

Table A.18 continued from previous page

Variedade	Predito	Grupo	Variedade	Predito	Grupo
REF_5	4,0%	g2			
VAR_2365	4,0%	g2			
VAR_2378	4,0%	g2			

Tabela A.19. Agrupamento final para conjunto de dados de mosaico para ensaio 2017 (cana soca)

Variedade	Predito	Grupo	Variedade	Predito	Grupo
REF_6	71,4%	g1	VAR_1844	7,7%	g4
REF_1	63,7%	g1	VAR_1810	6,3%	g4
REF_10	60,3%	g1	REF_8	5,7%	g4
VAR_2445	52,0%	g1	VAR_1719	4,2%	g4
REF_3	49,0%	g1	VAR_2436	4,1%	g4
VAR_2393	48,6%	g1	VAR_1795	3,8%	g4
VAR_2421	45,0%	g2	VAR_2367	2,8%	g4
REF_7	43,2%	g2	VAR_2405	2,8%	g4
VAR_2395	39,0%	g2	VAR_2044	2,8%	g4
VAR_1704	34,8%	g2	VAR_1832	1,8%	g4
VAR_2427	31,8%	g2	VAR_2402	1,4%	g4
REF_2	31,5%	g2	VAR_2388	1,4%	g4
VAR_2365	22,8%	g3	REF_5	1,4%	g4
VAR_2368	22,3%	g3	REF_4	1,4%	g4
VAR_1591	21,6%	g3			
VAR_2026	20,4%	g3			
VAR_2378	15,2%	g3			
VAR_1560	14,7%	g3			
VAR_2437	8,6%	g3			
VAR_2410	8,4%	g3			