

EVANDRO DE ÁVILA E LARA

**REGRESSÃO LOGÍSTICA POLITÔMICA ORDINAL: AVALIAÇÃO DO
POTENCIAL DE *Clonostachys rosea* NO BIOCONTROLE DE *Botrytis cinerea***

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2012

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

L318r
2012

Lara, Evandro de Ávila e, 1985-
Regressão logística politômica ordinal : avaliação do
potencial de *Clonostachys rosea* no biocontrole de
Botrytis cinerea / Evandro de Ávila e Lara. – Viçosa, MG,
2012.
xi, 39f. : il. ; 29cm.

Orientador: Sebastião Martins Filho.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 38-39.

1. Modelos matemáticos. 2. Modelos lineares (Estatística).
3. Banco de dados. 4. Tomate - Doenças e pragas. 5. Morango
- Doenças e pragas. I. Universidade Federal de Viçosa.
Departamento de Estatística. Programa de Pós-Graduação em
Estatística Aplicada e Biometria. II. Título.

CDD 22. ed. 519.5

EVANDRO DE ÁVILA E LARA

**REGRESSÃO LOGÍSTICA POLITÔMICA ORDINAL: AVALIAÇÃO DO
POTENCIAL DE *Clonostachys rosea* NO BIOCONTROLE DE *Botrytis cinerea***

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 23 de julho de 2012.

Leonardo Lopes Bhering

Gérson Rodrigues dos Santos

Luis Antônio Maffia
(Coorientador)

Carlos Henrique Osório Silva
(Coorientador)

Sebastião Martins Filho
(Orientador)

Aos meus pais Márcio Roberto Lara e Dalva Dulce de Ávila Lara, pelos conselhos seguros, esforços jamais negados e pelos exemplos oferecidos.

Aos meus irmãos Leandro de Ávila Lara e Márcia de Ávila Lara, pelos exemplos de vontade e dedicação.

À minha namorada Regiane Maria Soares Ramos, pelos momentos de confiança, amor e nos momentos que estive ausente, compreensão.

Dedico.

“Eu vou chegar, pedir e agradecer, pois, às vezes, a vitória de um homem se esconde num gesto forte em que só ele pode ver.”

Marcelo Falcão Custódio

AGRADECIMENTOS

A Deus pelas oportunidades e pela força diária. Em especial, nesse momento.

À Universidade Federal de Viçosa, por intermédio do Programa de Mestrado em Estatística Aplicada e Biometria, pela oportunidade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pelo apoio ao desenvolvimento deste projeto de pesquisa.

Ao professor Sebastião Martins Filho, pelo apoio, paciência, sabedoria e aprendizado durante este tempo que trabalhamos juntos.

A todos os professores e funcionários do Departamento de Estatística da Universidade Federal de Viçosa que contribuíram para minha formação profissional, sobretudo professores José Ivo Ribeiro Junior, Carlos Henrique Osório Silva, Fabyano Fonseca e Silva, Luiz Alexandre Peternelli, Antônio Policarpo Carneiro e Gérson Rodrigues dos Santos.

Ao professor Luis Antônio Maffia que, gentilmente, nos cedeu os dados e ao professor Leonardo Lopes Bhering pela revisão e orientações.

Aos colegas do mestrado do semestre 2010/1.

A todos que de alguma forma contribuíram, obrigado.

BIOGRAFIA

EVANDRO DE ÁVILA E LARA, filho de Dalva Dulce de Ávila Lara e Márcio Roberto Lara, nasceu em Três Pontas, Minas Gerais, no dia 24 de outubro de 1985.

Em março de 2004, iniciou o curso de graduação em Matemática na Universidade Federal de Viçosa – UFV, Minas Gerais, o qual concluiu em julho de 2009.

Em março de 2010, ingressou no programa de pós-graduação, em nível de mestrado, em Estatística Aplicada e Biometria, da Universidade Federal de Viçosa, submetendo-se a defesa em julho de 2012.

ÍNDICE

	Página
LISTA DE FIGURAS.....	VIII
LISTA DE TABELAS.....	IX
RESUMO.....	X
ABSTRACT.....	XI
1. INTRODUÇÃO.....	1
2. REVISÃO DE LITERATURA.....	2
2.1. Modelo de Regressão Logística Simples.....	3
2.2. Modelo de Regressão Logística Múltipla.....	5
2.2.1. Razão de chances.....	6
2.2.2. Estimação dos parâmetros.....	7
2.2.3. Teste da razão de verossimilhança.....	8
2.2.4. Teste Wald.....	10
2.3. Modelo de Regressão Logística Politômica.....	10
2.3.1. Modelo de Regressão Logística Politômica Ordinal.....	11
2.3.1.1. Teste score de chances proporcionais.....	11
2.3.1.2. Modelo de chances proporcionais.....	12
2.3.1.3. Modelo de chances proporcionais parciais.....	13
2.3.1.4. Modelo de chances não proporcionais.....	14
3. MATERIAL E MÉTODOS.....	17
3.1. Descrição do experimento.....	17
3.2. Modelo de chances proporcionais.....	18
3.2.1. Comandos do SAS para o modelo de chances proporcionais.....	19
3.3. Modelo de chances proporcionais parciais.....	20
3.3.1. Comandos do SAS para o modelo de chances proporcionais parciais.....	20

3.4. Modelo de chances não proporcionais.....	22
3.4.1. Comandos do SAS para o modelo de chances não proporcionais.....	22
3.5. Estimação dos Parâmetros e Adequação do Modelo.....	24
4. RESULTADOS.....	26
4.1. Modelo de chances proporcionais.....	26
4.2. Modelo de chances não proporcionais.....	27
4.3. Modelo de chances proporcionais parciais.....	30
4.4. Análise dos modelos de chances proporcionais parciais e não proporcionais.....	34
5. DISCUSSÃO.....	35
6. CONCLUSÃO.....	37
7. REFERÊNCIAS BIBLIOGRÁFICAS.....	38

LISTA DE FIGURAS

	Página
Figura 1 - Curva em S gerada por uma função logística e a relação linear entre uma única variável x e o <i>logit</i> $g(x)$	4

LISTA DE TABELAS

	Página
Tabela 1 – Variáveis dependente e independentes utilizadas no experimento no qual foi avaliado o potencial do <i>Clonostachys rosea</i> no biocontrole do Mofo Cinzento, causado por <i>Botrytis cinerea</i> , nas culturas de morango e tomate, (2010).....	17
Tabela 2 – Parâmetros do modelo de chances proporcionais do experimento no qual foi avaliado o potencial do <i>Clonostachys rosea</i> , no biocontrole do mofo cinzento, causado por <i>Botrytis cinerea</i> , nas culturas de morangueiro e tomateiro. Viçosa, 2010.....	26
Tabela 3 – Parâmetros do modelo de chances não proporcionais do experimento no qual foi avaliado o potencial do <i>Clonostachys rosea</i> , no biocontrole do mofo cinzento, causado por <i>Botrytis cinerea</i> , nas culturas de morangueiro e tomateiro. Viçosa, 2010.....	27
Tabela 4 – Contrastes realizados para investigação da pressuposição de chances proporcionais das variáveis explicativas.....	30
Tabela 5 – Contrastes realizados para investigação da pressuposição de chances proporcionais das variáveis explicativas já com a variável <i>Inoculação do B. cinerea</i> como uma variável proporcional.....	31
Tabela 6 – Parâmetros do modelo de chances proporcionais parciais do experimento no qual foi avaliado o potencial do <i>Clonostachys rosea</i> , no biocontrole do mofo cinzento, causado por <i>Botrytis cinerea</i> , nas culturas de morangueiro e tomateiro. Viçosa, 2010.....	32
Tabela 7 - Análise comparativa dos modelos de chances proporcionais parciais e não proporcionais do experimento no qual foi avaliado o potencial do <i>Clonostachys rosea</i> , no biocontrole do mofo cinzento, causado por <i>Botrytis cinerea</i> , nas culturas de morangueiro e tomateiro. Viçosa, 2010.....	34

RESUMO

LARA, Evandro de Ávila e, M.Sc., Universidade Federal de Viçosa, Julho de 2012. **Regressão logística politômica ordinal: Avaliação do potencial de *Clonostachys rosea* no biocontrole de *Botrytis cinerea*.** Orientador: Sebastião Martins Filho. Coorientadores: Carlos Henrique Osório Silva e Luiz Antônio Maffia.

O uso da regressão logística como uma ferramenta estatística para modelar a probabilidade de um evento em função de uma ou mais variáveis explicativas, tem crescido entre pesquisadores em várias áreas, inclusive na Fitopatologia. À respeito da regressão logística dicotômica, na qual a variável resposta é do tipo binária ou *dummy*, é extenso o número de trabalhos na literatura que abordam a modelagem, as pressuposições e a interpretação das análises, bem como alternativas de implementação em pacotes estatísticos. No entanto, quando a variável resposta requer que se utilize três ou mais categorias, o número de publicações é escasso. Isso devido não somente à escassez de publicações relevantes sobre o assunto, mas também à inerente dificuldade de abrangência sobre o tema. No presente trabalho aborda-se a aplicabilidade do modelo de regressão logística politômica ordinal, bem como as diferenças entre os modelos de chances proporcionais, chances proporcionais parciais e chances não proporcionais. Para isso, foram analisados dados de um experimento em que se avaliou o potencial do fungo antagonista *Clonostachys rosea* no biocontrole da doença denominada “mofo cinzento”, causada por *Botrytis cinerea* em morangueiro e tomateiro. Os modelos de chances proporcionais parciais e não proporcionais foram ajustados e comparados, uma vez que o teste score de proporcionalidade acusou rejeição da pressuposição de chances proporcionais. As estimativas dos coeficientes dos modelos bem como das razões de chances foram interpretadas em termos práticos para a Fitopatologia. A regressão logística politômica ordinal se apresentou como uma importante ferramenta estatística para predição de valores, mostrando o potencial do *C. rosea* em se tornar um produto comercial a ser desenvolvido e usado no controle biológico da doença, pois a aplicação de *C. rosea* foi tão ou mais eficiente do que a utilização de fungicidas no controle do mofo cinzento.

ABSTRACT

LARA, Evandro de Avila e, M.Sc., Universidade Federal de Viçosa, July, 2012. **Polytomous ordinal logistic regression: Assessing the potential of *Clonostachys rosea* in biocontrol of *Botrytis cinerea*.** Adviser: Sebastião Martins Filho. Co-Advisers: Carlos Henrique Osorio Silva and Luiz Antonio Maffia.

The use of logistic regression modeling as a tool for modeling statistical probability of an event as a function of one or more independent variables, has grown among researchers in several areas, including Phytopathology. At about the dichotomous logistic regression in which the dependent variable is the type binary or dummy, is the extensive number of studies in the literature that discuss the modeling assumptions and the interpretation of the analyzes, as well as alternatives for implementation in statistical packages. However, when the variable response requires the use three or more categories, the number of publications is scarce. This is not only due to the scarcity of relevant publications on the subject, but also the inherent difficulty of coverage on the subject. In this paper we address the applicability of the model polytomous ordinal logistic regression, as well as differences between the proportional odds models, nonproportional and partial proportional odds. For this, we analyzed data from an experiment in which we evaluated the potential antagonistic fungus *Clonostachys rosea* in biocontrol of the disease called "gray mold", caused by *Botrytis cinerea* in strawberry and tomato. The partial proportional odds models and nonproportional were adjusted and compared, since the proportionality test score accused rejection of the proportional odds assumption. The estimates of the model coefficients as well as the odds ratios were interpreted in practical terms for Phytopathology. The polytomous ordinal logistic regression is introduced as an important statistical tool for predicting values, showing the potential of *C. rosea* in becoming a commercial product to be developed and used in the biological control of the disease, because the application of *C. rosea* was as or more effective than the use of fungicides in the control of gray mold.

1. INTRODUÇÃO

A regressão logística é uma técnica estatística amplamente utilizada em diversas áreas do conhecimento, tais como: na medicina, em estudos epidemiológicos para avaliar fatores de riscos na incidência de doenças, na ciência social, para explicar as intenções de voto em atos eleitorais e na econometria, na predição de grupos de risco para obtenção de crédito. Essa regressão tem, na sua praticidade de interpretação dos resultados, seu grande diferencial. Esta técnica apresenta ainda qualidade ao ser capaz, não somente de mostrar qual combinação de variáveis explicativas é melhor, como também estimar sua significância na variável resposta.

Hosmer e Lemeshow (2002) citam que a regressão logística é usualmente utilizada quando se tem o interesse na avaliação da influência de fatores sobre uma resposta dicotômica ou politômica. Quando a variável resposta é agrupada em mais que duas categorias ela é denominada politômica.

O número de trabalhos na literatura com o uso da regressão logística dicotômica foram impulsionados após os trabalhos de Cox (1970), Agresti (1990), Hosmer e Lemeshow (1989), entre outros. Entretanto, essa abundância se contrasta com a escassez de trabalhos com o uso das variáveis respostas politômicas. Esta escassez pode ser atribuída não somente à maior complexidade do modelo, mas também devido à dificuldade de validação dos seus pressupostos bem como às reduzidas opções de modelagem oferecidas em softwares estatísticos comerciais (ABREU, 2009). Assim, há carência de materiais teóricos e práticos que abordam regressão logística quando a variável dependente é do tipo politômica, especialmente nas ciências agrárias.

Em vista desta carência, este trabalho objetivou detalhar a aplicabilidade do modelo de regressão logística politômica ordinal, bem como as diferenças entre os modelos de chances proporcionais, chances proporcionais parciais e chances não proporcionais, com a utilização de dados reais obtidos no departamento de Fitopatologia da UFV, oriundos de um experimento no qual foi avaliado o potencial do fungo *Clonostachys rosea* no controle biológico e no manejo integrado da doença denominada “mofo cinzento”, causada pelo fungo *Botrytis cinerea*, em morangueiro e tomateiro.

2. REVISÃO DE LITERATURA

O termo regressão foi proposto por Francis Galton em 1885, em um estudo que demonstrou que a altura dos filhos não tende a refletir a altura dos pais, mas tende a regredir para a média da população (Johnson & Wichern, 1992). De acordo com Ryan (2009), a análise de regressão é uma das técnicas estatísticas mais utilizadas e seu uso está presente em quase todos os campos de aplicação. Weisberg (2005) afirma que a análise de regressão é a parte central de muitos projetos de pesquisa. No entanto, Allison (1999) já havia reportado pesquisadores utilizando inadequadamente a técnica de regressão linear para a análise de variáveis respostas qualitativas nominal e ordinal, o que, em sua opinião, se deve ao desconhecimento de técnicas mais avançadas. Essencialmente, o termo análise de regressão significa ajustar um modelo que relacione uma variável resposta (ou dependente) a uma ou mais variáveis explicativas (ou independentes). Não é o objetivo deste texto abordar as inúmeras alternativas de modelos que podem ser empregados em uma análise de regressão. Nosso objetivo é uma abordagem dos modelos de regressão logística politômica ordinal, nos quais a variável resposta é do tipo categórica.

Uma variável resposta é definida como sendo do tipo categórica quando ela possui uma escala de medida formada por um conjunto de categorias (AGRESTI, 1990). As categorias que a variável dependente assume podem ser de natureza nominal ou ordinal. Quando há uma ordem natural entre as categorias, o contexto é de regressão logística ordinal, caso contrário, assume-se o contexto da regressão logística nominal.

À respeito das variáveis independentes ou fatores, o número de fatores também é de extrema importância para a escolha do modelo. Isto porque, caso o experimento leve em consideração dois ou mais fatores, a regressão se torna uma regressão logística múltipla. Portanto, pode-se dizer que a regressão logística múltipla é uma generalização da regressão logística simples (com apenas um fator independente). Atenção especial deve ser dada para não confundir os termos múltipla com politômica. Enquanto o termo múltipla refere-se à múltiplas variáveis independentes, o termo politômica ou multinomial refere-se a múltiplas categorias de resposta da variável dependente.

2.1. MODELO DE REGRESSÃO LOGÍSTICA SIMPLES

É difícil precisar exatamente quando se usou o modelo logístico pela primeira vez, porém Cox e Snell (1989) e Hosmer e Lemeshow (2002) concordam que o modelo de regressão logística ganhou reconhecimento após o trabalho de Truett et al., (1967) que analisava o risco de doença coronariana. Esse trabalho até hoje é considerado um marco inicial dos estudos envolvendo regressão logística nas áreas da saúde. McLachlan (1992) também afirma que as primeiras aplicações do modelo logístico foram em estudos prospectivos de doenças coronarianas. Contudo, nessas aplicações, os autores realizaram o processo de estimação de parâmetros sob a suposição de normalidade, que se torna desnecessária quando a estimação é feita por máxima verossimilhança via métodos numéricos. O procedimento de estimação em um contexto mais genérico foi proposto por Day e Kerridge (1967) e por Walker e Duncan (1967).

Hosmer e Lemeshow (2002) citam duas principais justificativas para adoção do modelo logístico na análise de variáveis resposta dicotômicas: sob o ponto de vista matemático, é razoavelmente flexível e fácil de ser utilizado; permite uma interpretação de resultados bastante rica e direta.

O modelo de regressão logística simples permite estimar a probabilidade de um evento, indicado por uma variável Y , sendo $Y=1$ quando o evento ocorre e $Y=0$ quando não ocorre. O modelo é do tipo simples pois inclui apenas uma variável explicativa X , sendo esta em geral do tipo quantitativa, mas podendo ser também do tipo categórica nominal ou ordinal. O modelo é dado por :

$$P(Y = 1 | X = x) = \frac{\exp [\beta_0 + \beta x]}{1 + \exp [\beta_0 + \beta x]} + e, \quad \text{em que}$$

- ✓ β_0 : parâmetro de intercepto,
- ✓ β : parâmetro da regressão,
- ✓ $Y = 1$: indica a ocorrência do evento,
- ✓ X : variável explicativa (ou fator de risco),
- ✓ e : erro aleatório do modelo

Consequentemente, $P(Y=0 | X = x) = \frac{1}{1 + \exp [\beta_0 + \beta x]} + e$. Assim, a função logito, definida como o logaritmo natural da chance do evento, resulta no modelo linear,

$$\text{logit} \left(\frac{P(Y = 1 | X = x)}{P(Y = 0 | X = x)} \right) = \ln \left(\frac{P(Y = 1 | X = x)}{P(Y = 0 | X = x)} \right) = \beta_0 + \beta x .$$

A Figura 1 apresenta a função logística com o seu formato característico em ‘S’ e a relação linear entre uma única variável x e o *logit* g(x).

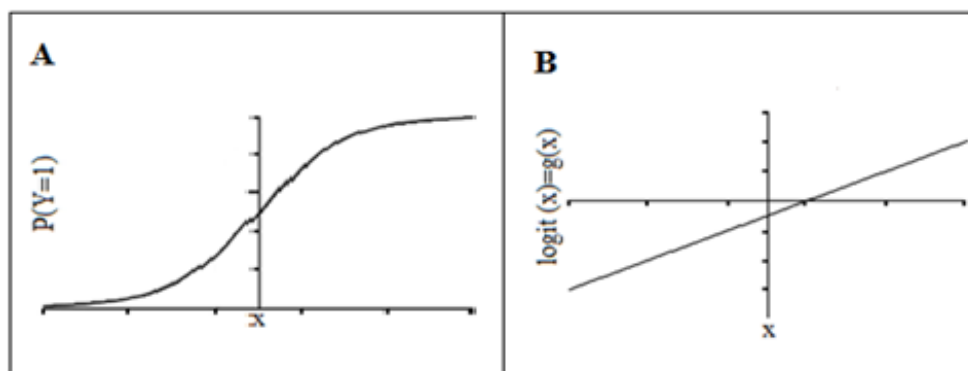


Figura 1. Curva em S gerada por uma função logística (A) e a relação linear entre uma única variável x e o *logit* g(x) (B).

Quando a variável resposta for politômica com L categorias de respostas, o modelo anterior é ajustado tratando-se duas categorias por vez e serão formados L-1 logitos :

$$\text{logit} (Y = k) = \beta_{0k} + \beta_k x \text{ em que } k = 1, 2, 3 \dots L-1 .$$

Assim, enquanto estima-se um único parâmetro para o intercepto no caso de variável resposta dicotômica, L-1 parâmetros são estimados para o intercepto quando a variável resposta é politômica. Detalhes da estimação dos parâmetros por máxima verossimilhança serão apresentados adiante no texto.

2.2. MODELO DE REGRESSÃO LOGÍSTICA MÚLTIPLA

Hosmer e Lemeshow (2002) estabeleceram a generalização da regressão logística simples para a múltipla da seguinte forma: Seja um conjunto com p variáveis independentes ou explicativas e denote por $\mathbf{X}_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip})$ o vetor da i -ésima linha da correspondente matriz \mathbf{X} formada pelas variáveis explicativas; denote por $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \dots, \beta_p)^T$, o vetor de parâmetros desconhecidos associados às p variáveis explicativas. No modelo de regressão múltipla dicotômica, a probabilidade de sucesso é dada por:

$$\pi(\mathbf{X}_i) = P(Y=1|\mathbf{X}=\mathbf{X}_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})} = \frac{\exp(\beta_0 + X_i \boldsymbol{\beta})}{1 + \exp(\beta_0 + X_i \boldsymbol{\beta})} \quad (1)$$

E o complemento desta probabilidade se torna:

$$1 - \pi(\mathbf{X}_i) = P(Y=0|\mathbf{X}=\mathbf{X}_i) = \frac{1}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})} = \frac{1}{1 + \exp(\beta_0 + X_i \boldsymbol{\beta})} \quad (2)$$

$$\text{Note que } \frac{\pi(\mathbf{X}_i)}{1 - \pi(\mathbf{X}_i)} = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \quad (3)$$

Ao aplicar o logaritmo natural em (3) tem-se a função de ligação $g(x)$ que, no contexto de Modelos Lineares Generalizados, é uma função monótona, diferenciável e que relaciona a média de Y , dada por $E(Y) = P(Y=1)$, ao seu preditor linear (Hosmer e Lemeshow, 2002) dada por:

$$g(\mathbf{X}_i) = \ln(\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})) \quad \text{ou} \\ g(\mathbf{X}_i) = (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) = \beta_0 + X_i \boldsymbol{\beta} \quad (4)$$

O modelo (1) é denominado regressão logística múltipla, já que vem de uma transformação logística, também conhecida como transformação *logit*, com múltiplos parâmetros. Estimação dos parâmetros e inferências de interesse serão abordados adiante no texto.

Deve ser notada a ausência do erro aleatório ξ no modelo apresentado (1) já que o lado esquerdo do modelo é $E(Y|\mathbf{X}_i)$ em que Y segue uma Bernoulli $(1, \pi(x))$. Porém, podemos escrever:

$$P(Y=1 | \mathbf{X} = \mathbf{x}_i) = E(Y|\mathbf{x}_i) = \pi(\mathbf{x}_i) = P(Y=1|\mathbf{x}_i) = \frac{\exp g(\mathbf{x}_i)}{1 + \exp g(\mathbf{x}_i)} + \xi_i \quad (5)$$

em que ξ assume dois valores. Se $Y=1$, $\xi = 1-\pi(x)$ com probabilidade $\pi(x)$, se $Y = 0$, $\xi = -\pi(x)$ com probabilidade $1 - \pi(x)$. Assim, os erros não são normais e não há a homogeneidade da variância pois para todo i , $i=\{1, 2, \dots, n\}$

$$i) E(\xi_i | x_i) = 0$$

$$ii) V(\xi_i | x_i) = [\pi(x_i) / (1 - \pi(x_i))]$$

$$iii) Cov(\xi_i, \xi_j) = 0, \text{ se } i \neq j$$

2.2.1. RAZÃO DE CHANCES

A razão de chances (OR *do inglês Odds Ratio*) é a estatística mais comumente utilizada para inferir em termos de probabilidades relacionadas ao evento e às variáveis de interesse. Ela é definida como sendo a exponencial da razão entre as chances de ocorrência de dois eventos (Hosmer e Lemeshow, 2002). Por definição, chance de um evento é o quociente entre a probabilidade do evento ocorrer pela probabilidade do mesmo evento não ocorrer. Assim, sendo p e q as probabilidades de ocorrência de dois eventos, então a razão de chances entre os eventos é:

$$OR = \exp\left(\frac{p}{(1-p)} / \frac{q}{(1-q)}\right)$$

Pela equação (3) :

$$\ln\left[\frac{\pi(x_{i+1})}{1-\pi(x_{i+1})}\right] - \ln\left[\frac{\pi(x_i)}{1-\pi(x_i)}\right] = \beta_0 + \beta_1(X_{i+1}) - \beta_0 - \beta_1(X_i) = \beta_1(X_{i+1} - X_i)$$

Admitindo X uma variável quantitativa e fazendo $X_{i+1} - X_i = 1$, tem-se que $\beta_1(X_{i+1} - X_i) = \beta_1$

Desta forma, β_1 representa o incremento no valor do logaritmo da chance, dada pela expressão (3), devido ao aumento de uma unidade em x .

Aplicando-se exponencial em (5), temos que:

$$\frac{\pi(x_i + 1)}{1 - \pi(x_i + 1)} / \frac{\pi(x_i)}{1 - \pi(x_i)} = \exp\beta_1 \quad (7)$$

Assim, a expressão (6) é chamada de razão das chances ou *Odds Ratio* e compara a probabilidade de sucesso para $x + 1$ com a probabilidade de sucesso para x . Note que, admitindo X como uma variável categórica, tem-se

$$\ln\left[\frac{\pi(1)}{1-\pi(1)}\right] - \ln\left[\frac{\pi(0)}{1-\pi(0)}\right] = \beta_0 + \beta_1(1) - \beta_0 - \beta_1(0) = \beta_1$$

2.2.2. ESTIMAÇÃO DOS PARÂMETROS POR MÁXIMA VEROSSIMILHANÇA

O método de estimação de parâmetros mais comumente usado na regressão logística é o da máxima verossimilhança. Detalhes da metodologia são apresentados em diversas referências, tais como em (RYAN, 2009), do qual foram adaptados os resultados apresentados a seguir. A função de probabilidade de Y para a regressão logística binária simples é:

$$f(y_i, \pi_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad \text{com } y_i = \{0,1\} \quad (8)$$

Assim, a função de verossimilhança será dado por:

$$L(\beta) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (9)$$

O método de máxima verossimilhança consiste em estimar β que maximiza $L(\beta)$. Aplicando-se o logaritmo natural em ambos os lados da equação (9) facilita a tarefa.

$$\begin{aligned} l(\beta) &= \ln [L(\beta)] = \ln [\prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i}] \\ &= \sum_{i=1}^n [y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)] \\ &= \sum_{i=1}^n [y_i \ln\left(\frac{\pi_i}{1 - \pi_i}\right) + \ln(1 - \pi_i)]. \end{aligned} \quad (10)$$

Substituindo (2) e (3) em (10), temos

$$\begin{aligned} l(\beta) &= \sum_{i=1}^n [y_i (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) + \ln\left(\frac{1}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}\right)] \\ l(\beta) &= \sum_{i=1}^n [y_i (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) - \ln(1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}))] \end{aligned} \quad (11)$$

Os valores de β_0, \dots, β_p que maximizam essa função são obtidos igualando-se as $p + 1$ equações a zero e esse máximo é garantido pois $0 < P [Y_i = y_i | x_i] < 1$ já que a função logaritmo é estritamente crescente.

$$\frac{\partial l(\beta)}{\partial \beta_0} = \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\pi}_i = 0 \quad (12)$$

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^n x_{ij} y_i - \sum_{i=1}^n x_{ij} \hat{\pi}_i = 0 \quad \text{para } j = \{1,2,3,\dots,p\} \quad (13)$$

Tais equações são não lineares nos parâmetros e requerem o emprego de processos iterativos como Newton-Raphson e escore de Fisher na sua solução (Gourieroux e Monfort, 1995). Este método é definido expandindo-se a função $U(\beta)$ em torno do ponto inicial $\beta^{(0)}$, tal que

$$U(\beta) \approx U(\beta^{(0)}) + U'(\beta^{(0)}) (\beta - \beta^{(0)})$$

sendo $U(\beta)$ as derivadas de primeira ordem do logaritmo da função de verossimilhança em relação aos parâmetros do modelo e $U'(\beta^{(0)})$ são as derivadas de segunda ordem do logaritmo da função de verossimilhança. Ao repetir o processo descrito acima, chega-se ao processo iterativo até que o valor de β se estabilize tal que

$$\beta^{(m+1)} = \beta^{(m)} + [-U'(\beta^{(m)})]^{-1} U'(\beta^{(m)}), \text{ sendo } m=0,1,\dots$$

Como a matriz $-U'(\beta)$ pode não ser positiva definida e portanto não invertível, ela é substituída pela matriz de informação de Fisher (I). Assim,

$$\beta^{(m+1)} = \beta^{(m)} + [-I(\beta^{(m)})]^{-1} U'(\beta^{(m)}) \text{ em que}$$

$$I(\widehat{\beta}) = \begin{pmatrix} \frac{\partial^2 \ln L(\beta_0, \beta_1)}{\partial \beta_0^2} & \frac{\partial^2 \ln L(\beta_0, \beta_1)}{\partial \beta_0 \beta_1} \\ \frac{\partial^2 \ln L(\beta_0, \beta_1)}{\partial \beta_0 \beta_1} & \frac{\partial^2 \ln L(\beta_0, \beta_1)}{\partial \beta_1^2} \end{pmatrix}.$$

Para o modelo de regressão logística múltiplo, o processo se torna análogo ao descrito acima usado na estimação do modelo de regressão logística simples. Podem-se obter maiores informações em Gourieroux e Monfort (1995).

2.2.3 – TESTE DA RAZÃO DE VEROSSIMILHANÇA (TRV)

Após a estimação dos parâmetros do modelo, é iniciado a fase da verificação do ajuste do modelo aos dados. O teste da razão de verossimilhança consiste em comparar as probabilidades conjuntas da amostra sob dois cenários: com ou sem determinadas variáveis incluídas no modelo. O caso mais extremo, que é o de modelo inadequado, tem-se a hipótese nula de que os K parâmetros das variáveis

explicativas sejam iguais a zero. Neste caso o TRV emprega o modelo saturado para essa comparação, modelo este em que o número de parâmetros é igual ao número de observações.

Para essa comparação é feita de acordo com a seguinte estatística D:

$$D = - 2 \ln \left[\frac{\text{Verossimilhança do modelo ajustado}}{\text{Verossimilhança do modelo saturado}} \right] \quad (14)$$

Aplicando a equação (11) em (14), temos

$$D = - 2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}_i}{1 - y_i} \right) \right] \quad (15)$$

A Estatística D, chamada de *Deviance*, desempenha um papel fundamental em algumas aproximações para verificar a eficácia do ajuste. Sob a hipótese de nulidade, prova-se que em grandes amostras (teste assintótico) a estatística D segue uma distribuição de qui-quadrado com p graus de liberdade (Hosmer e Lemeshow, 2002).

No modelo saturado, no entanto, $\hat{\pi}_i = y_i$ e a verossimilhança nesse caso vale 1. Assim, a equação (14) se torna

$$D = - 2 \ln [\text{Verossimilhança do modelo ajustado}] \quad (16)$$

Com o propósito de verificar a significância de uma ou mais variável (eis) independente(s) em um modelo ajustado, a comparação do modelo com e sem a(s) variável(eis) em questão se dá por meio da comparação da estatística *Deviance* dos respectivos modelos. Logo,

$$G = D (\text{modelo sem as variáveis}) - D (\text{modelo com as variáveis}) \quad (17)$$

Ou então, aplicando (16) em (17)

$$G = - 2 \ln \left[\frac{\text{verossimilhança sem a(s) variável (eis)}}{\text{verossimilhança com a(s) variável (eis)}} \right] \quad (18)$$

Assim, a estatística D, sob a hipótese de que o modelo ajustado é correto, tem distribuição assintótica χ^2 com $p = K (J - 2)$ graus de liberdade, sendo K o número de variáveis explicativas e J o número de categorias da variável resposta presentes no modelo.

2.2.4 – TESTE WALD

O teste Wald é outro teste bastante adotado para testar hipóteses relativas a um único parâmetro β_j , para $j = 1, 2, \dots, p$. Ele é baseado na distribuição assintótica de $\hat{\beta}$ e é uma generalização do teste t de Student (Wald, 1943). Porém, Hauck e Donner (1977) examinaram o desempenho do teste de Wald e descobriram que ele se comporta de maneira estranha, em determinadas situações; frequentemente não rejeitando a hipótese nula quando o coeficiente é significativo. Eles recomendam a utilização do teste da razão de verossimilhança para testar se realmente o coeficiente não é significativo quando o teste de Wald não rejeita a hipótese nula.

Considerando-se a hipótese nula $H_0: \beta = \beta_0$, a estatística para este teste é dada por:

$$W = \frac{(\hat{\beta} - \beta_0)^2}{\widehat{Var}(\hat{\beta})}, \quad (19)$$

que sob H_0 , tem aproximadamente uma distribuição qui-quadrado com p graus de liberdade.

2.3 – MODELO DE REGRESSÃO LOGÍSTICA POLITÔMICA

O modelo de regressão logística, originalmente desenvolvido para variáveis resposta binárias, é extensível para variáveis resposta politômicas. Este modelo surgiu como generalização de uma variável discreta em três ou mais categorias.

De acordo com Agresti (2007), modelos *logit* com multicategorias utilizam todos os pares de categorias para especificar a chance da resposta (ou do evento) ocorrer em uma categoria em relação a outra categoria.

Dentre as categorias da variável discreta, uma deve ser considerada como a de referência. Dessa forma, se a j -ésima categoria de resposta é usada para como referência, o *logit* para esse modelo é

$$\ln\left(\frac{\pi_k}{\pi_j}\right), k = 1, 2, 3, \dots, J-1 \quad \text{em que } \hat{\pi}_j = \frac{\exp(\hat{\mu}_j)}{\sum_{h=1}^{J-1} \exp(\hat{\mu}_h)}$$

$$\text{Assim, temos que } \ln\left(\frac{\pi_k}{\pi_j}\right) = \beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2 + \dots + \beta_{pj}x_p + \varepsilon$$

2.3.1.REGRESSÃO LOGÍSTICA POLITÔMICA ORDINAL

O modelo de regressão logística ordinal vem sendo amplamente usado por se adequar melhor a dados que possuem respostas qualitativas e que podem ser categorizadas seguindo uma ordenação natural tais como a altura média de uma população (baixa, média e alta), o nível de escolaridade de um trabalhador (fundamental, médio e superior) ou até mesmo o estágio de uma doença (inicial, intermediário e terminal).

São vários os modelos ordinais, o modelo de chances proporcionais (POM), o modelo de chances proporcionais parciais (PPOM) e o modelo de chances não proporcionais (NPOM). Com maior destaque para o modelo de chances proporcionais, este vem sendo usado na literatura sem uma constatação básica e importante sobre suas pressuposições (Abreu,2009). Este modelo, descrito por McCullagh (1980), assume como pressuposição básica que os coeficientes, para cada variável explicativa do modelo, são os mesmos para todos os logitos. Assim, essa pressuposição deve ser testada antes do ajuste para esse modelo.

2.3.1.1 – TESTE SCORE PARA CHANCES PROPORCIONAIS

Esse teste é obtido a partir da função score, sendo a estatística do teste dada por(Corrar,2007):

$$S = U'(\theta_0) [-F(\theta_0)]^{-1} U(\theta_0), \quad (20)$$

em que $U(\theta_0)$ é a função score $U(\theta_0) = \frac{\partial \log L(\theta)}{\partial \theta}$ avaliada em θ_0 , $[-F(\theta_0)]^{-1}$ é a inversa de $-F(\theta_0)$, a matriz de variâncias e covariâncias observada de $U(\theta)$, também avaliada em θ_0 e $U'(\theta_0)$ representa a derivada da função score. Essa estatística segue assintoticamente (Corrar,2007) uma distribuição qui-quadrado com $p = K(J - 2)$ graus de liberdade, sendo K o número de variáveis explicativas e J o número de variáveis respostas presentes no modelo. Para amostras grandes, a hipótese H_0 de que o modelo é o de chances proporcionais, é rejeitada, a um nível α de significância e p graus de liberdade se $S > \chi^2_{(1-\alpha, p)}$. Um teste não significativo indica evidências que o modelo a ser ajustado é o de chances proporcionais.

2.3.1.2 – MODELO DE CHANCES PROPORCIONAIS

Para o modelo ordinal de chances proporcionais, Hosmer e Lemeshow (2002) dizem que a estratégia a buscar é a dicotomização das categorias da variável resposta. Assim, considerando um modelo politômico multinomial com três respostas a ser predito por quatro variáveis h, i, j e k , as probabilidades individuais de cada resposta são dadas por:

$$\begin{aligned}\pi_{hijk1} &= P(Y = 1 | X_1 = x_h, X_2 = x_i, X_3 = x_j, X_4 = x_k) \\ &= \frac{\exp(\beta_{01} + \beta_1 x_h + \beta_2 x_i + \beta_3 x_j + \beta_4 x_k)}{1 + \exp(\beta_{01} + \beta_1 x_h + \beta_2 x_i + \beta_3 x_j + \beta_4 x_k)}\end{aligned}\quad (21)$$

$$\begin{aligned}\pi_{hijk2} &= P(Y = 2 | X_1 = x_h, X_2 = x_i, X_3 = x_j, X_4 = x_k) \\ &= \frac{\exp(\beta_{02} + \beta_1 x_h + \beta_2 x_i + \beta_3 x_j + \beta_4 x_k)}{1 + \exp(\beta_{02} + \beta_1 x_h + \beta_2 x_i + \beta_3 x_j + \beta_4 x_k)}\end{aligned}\quad (22)$$

$$\begin{aligned}\pi_{hijk3} &= P(Y = 3 | X_1 = x_h, X_2 = x_i, X_3 = x_j, X_4 = x_k) \\ &= \frac{\exp(\beta_{03} + \beta_1 x_h + \beta_2 x_i + \beta_3 x_j + \beta_4 x_k)}{1 + \exp(\beta_{03} + \beta_1 x_h + \beta_2 x_i + \beta_3 x_j + \beta_4 x_k)}\end{aligned}\quad (23)$$

A análise dos dados ocorre de modo análogo a uma variável dicotômica e é realizada combinando-se duas das categorias de resposta por vez. Ou seja, faremos uma probabilidade acumulada tal que:

$$\theta_{hijk1} = \pi_{hijk1} \quad (24)$$

$$\theta_{hijk2} = \pi_{hijk1} + \pi_{hijk2} \quad (25)$$

$$\text{É fácil perceber que assim, } \pi_{hijk3} = 1 - \theta_{hijk2} \quad (26)$$

Note que para uma resposta ordinal com $L > 2$ categorias $L - 1$ logits cumulativos, baseados nas probabilidades acumuladas, são consideradas para cada subpopulação. Para três categorias da variável resposta, temos

$$\textit{logit}_1(\theta_{hijk1}) = \ln \frac{\pi_{hijk1}}{\pi_{hijk2} + \pi_{hijk3}} \quad \text{e} \quad \textit{logit}_2(\theta_{hijk2}) = \ln \frac{\pi_{hijk1} + \pi_{hijk2}}{\pi_{hijk3}}$$

$$\text{ou então } \textit{logit}_1(\theta_{hijk1}) = \beta_{01} + \beta'_L X_{hijk} \quad \text{e} \quad \textit{logit}_2(\theta_{hijk2}) = \beta_{02} + \beta'_L X_{hijk} \quad (27)$$

Assim, o modelo de chances (*odds*) proporcionais que leva em consideração ambas as *odds* citadas simultaneamente, para cada combinação das variáveis explicativas do modelo é:

$$\text{logit}(\theta_{hijkL}) = \beta_{0L} + \beta'_L X_{hijk} \text{ em que } L \text{ indexa os dois logitos.} \quad (28)$$

O parâmetro β_{01} representa o intercepto do primeiro logito cumulativo e o parâmetro β_{02} representa o intercepto do segundo logito cumulativo.

β'_L é um vetor de parâmetros dos efeitos para as variáveis explicativas.

X_{hijk} é um vetor representativo das variáveis explicativas.

É importante perceber que teremos $L - 1$ valores estimados para os interceptos e apenas um valor para cada variável explicativa do modelo, sendo este valor usado em todos os logitos.

2.3.1.3 – MODELO DE CHANCES PROPORCIONAIS PARCIAIS

Este modelo é usado quando a pressuposição de chances proporcionais é rejeitada para algumas variáveis explicativas, por meio do teste escore para proporcionalidade, o que na prática ocorre com bastante frequência (Hosmer e Lemeshow, 2002).

Nele é também ajustado um modelo de chances proporcionais. No entanto, para as variáveis onde essa pressuposição não é satisfeita, há um incremento de um coeficiente (γ_i), que é o efeito de não proporcionalidade associado em cada L -ésimo logito cumulativo. Assim, cada logito é ajustado associando este efeito às variáveis não proporcionais com as demais variáveis explicativas em que a pressuposição foi satisfeita. O modelo para o L -ésimo logito com as variáveis h e j não satisfazendo a pressuposição seria

$$\text{logit}(\theta_{hijkL}) = \beta_{0L} + (\beta_1 + \gamma_1) x_h + \beta_2 x_i + (\beta_3 + \gamma_3) x_h + \beta_4 x_k \quad (29)$$

2.3.1.4 – MODELO DE CHANCES NÃO PROPORCIONAIS

Uma outra alternativa à rejeição da pressuposição de chances proporcionais é ajustar o modelo de chances não proporcionais. Enquanto que no modelo de chances proporcionais, estima-se vários interceptos e um único parâmetro para cada variável independente o qual se admite como válido em todo logito, neste modelo são estimados vários interceptos bem como vários parâmetros para cada variável preditora em um respectivo logito (Hosmer e Lemeshow, 2002). Portanto, por exemplo, os modelos para os L-ésimo e M-ésimo logitos ajustados para as variáveis explicativas X_h, X_i, X_j e X_k , seriam:

$$\text{logit}(\theta_{h i j k L}) = \beta_{0L} + \beta_{1L} X_h + \beta_{2L} X_i + \beta_{3L} X_j + \beta_{4L} X_k \quad \text{e} \quad (30)$$

$$\text{logit}(\theta_{h i j k M}) = \beta_{0M} + \beta_{1M} X_h + \beta_{2M} X_i + \beta_{3M} X_j + \beta_{4M} X_k .$$

Uma vez que as variáveis explicativas não são proporcionais, ajusta-se um modelo para os logitos generalizados e não para os logitos cumulativos em que cada logito generalizado é formado com a probabilidade de cada categoria resposta sobre a categoria resposta de referência, escolhida de acordo com o interesse do pesquisador. Assim, num experimento com duas variáveis independentes h e i e quatro categorias de variáveis respostas, com

π_{hi1} = Probabilidade da resposta ocorrer na primeira categoria da variável dadas as condições h e i .

π_{hi2} = Probabilidade da resposta ocorrer na segunda categoria da variável dadas as condições h e i .

π_{hi3} = Probabilidade da resposta ocorrer na terceira categoria da variável dadas as condições h e i .

π_{hi4} = Probabilidade da resposta ocorrer na quarta categoria da variável dadas as condições h e i .

serão formados três logitos generalizados. Note que caso possuem L categorias da variável resposta serão formados $L-1$ logitos generalizados.

Tomado a primeira variável resposta como referência, tem-se

$$\text{logit}_{hi1} = \log \frac{\pi_{hi2}}{\pi_{hi1}} \quad \text{logit}_{hi2} = \log \frac{\pi_{hi3}}{\pi_{hi1}} \quad \text{logit}_{hi3} = \log \frac{\pi_{hi4}}{\pi_{hi1}} \quad (31)$$

Desse modo, o modelo a ser ajustado para os logitos generalizados é expresso por:

$$\text{logit}_{hik} = \beta_{0k} + \beta'_k \mathbf{X}_{hi} \text{ em que } k \text{ indexa os 3 logitos.} \quad (32)$$

Novamente, conforme explicado anteriormente no texto, a estimação dos parâmetros é pelo Método da Máxima Verossimilhança e a avaliação da adequação do modelo ajustado pelo Teste da Razão de Verossimilhança (TRV).

A construção da função de verossimilhança (Hosmer e Lemeshow, 2002) é ilustrada do seguinte modo:

Seja Y a variável dependente assumindo três categorias ditas Y_1 , Y_2 , e Y_3 . Com o auxílio de três variáveis binárias, atribui-se à categoria em que o valor pertence, o número 1 e às demais, o número 0.

Assim, se o valor observado pertence à segunda categoria, então $Y_1=0$, $Y_2=1$ e $Y_3=0$.

Logo, para uma amostra de n observações independentes, a função de verossimilhança é:

$$L(\beta) = \prod_{i=1}^n \pi_1(x_i)^{y_1} \pi_2(x_i)^{y_2} \pi_3(x_i)^{y_3} \quad (33)$$

Com a aplicação do logaritmo natural tem-se,

$$l(\beta) = \sum_{i=1}^n y_1 [\ln \pi_1(x_i)] + y_2 [\ln \pi_2(x_i)] + y_3 [\ln \pi_3(x_i)] \quad (34)$$

As equações de verossimilhança são obtidas tomando-se as derivadas parciais de $l(\beta)$ em relação a cada um dos parâmetros desconhecidos e igualando-as a zero.

$$\frac{\partial l(\beta)}{\partial \beta_{jk}} = \sum_{i=1}^n x_{ki} (y_{ji} - \pi_{ji}) \quad (35)$$

As equações acima são não lineares nos parâmetros e requerem o emprego de processos iterativos como Newton-Raphson e escore de Fisher (Gourieroux e Monfort, 1995) na sua solução já mencionados no texto.

Obtida as estimativas dos parâmetros β_j ($j = 0,1,2,\dots,p$) faz-se necessário avaliar a adequação do modelo ajustado comparando os valores observados da variável resposta com os valores preditos pelo modelo com, e sem, a covariável sob

investigação. Essa comparação é feita utilizando o Teste da Razão de Verossimilhança (TRV) expresso por:

$$\text{TRV} = - 2 \log \left[\frac{L_{SC}}{L_{CC}} \right] = - 2 [\log L_{SC} - \log L_{CC}] \quad (36)$$

e sob a hipótese nula de que os p coeficientes associados ao modelo são iguais a zero com distribuição Qui-quadrado e p graus de liberdade, pode-se avaliar o quanto da variação dos dados é explicada pela inclusão de termos no modelo a partir das diferenças de *Deviance*.

3. MATERIAL E MÉTODOS

Para esta aplicação, foram usados dados de um experimento realizado na Unidade de Controle Biológico do Departamento de Fitopatologia da Universidade Federal de Viçosa, no qual foi estudado o potencial do *Clonostachys rosea* no controle biológico do mofo cinzento, causado por *Botrytis cinerea*, nas culturas de morango e tomate.

O objetivo do trabalho, utilizando as variáveis explicativas *hospedeiro*, *dias*, *temperatura*, *inoculação do B. cinerea* e *concentração de C. rosea*, foi obter como variável resposta o número de conidióforos do *B. cinerea* presentes nas folhas das culturas de morango e tomate. Nas análises foi utilizado o software SAS (SAS INSTITUTE INC, 2008).

3.1. DESCRIÇÃO DO EXPERIMENTO

As variáveis usadas no modelo estão apresentadas na Tabela 1. Foram realizadas 512 observações, sendo 256 em morangueiros e 256 em tomateiros.

Tabela 1 - Variáveis utilizadas no experimento em que foi avaliado o potencial do *Clonostachys rosea*, no controle biológico do mofo cinzento nas culturas de morango e tomate, causado por *Botrytis cinerea*, Viçosa, 2010

Variável	Descrição	Valores assumidos
Y	Número de conidióforos de <i>B. cinerea</i> presentes na folha.	0 - Nenhum (0) 1 - Baixo (< 50) 2 - Alto (≥ 50)
X ₁	Hospedeiro	0 – Morangueiro 1 – Tomateiro
X ₂	Temperatura em que as plantas foram expostas após a inoculação do <i>B. cinerea</i> .	0 - 18°C 1 - 26°C
X ₃	Inoculação do <i>B. cinerea</i>	0 - Não Inoculou 1 – Inoculou
X ₄	Concentração de conidióforos de <i>C. rosea</i> aplicada nas folhas.	0 - 10 ⁴ 1 - 10 ⁷
X ₅	Número de dias em que se avaliou a variável resposta após a inoculação de <i>B. cinerea</i> .	2, 5, 8, 11, 14, 17, 20 e 23

3.2 – MODELO DE CHANCES PROPORCIONAIS

Após a categorização dos dados, foram obtidas as estimativas dos parâmetros do modelo de chances proporcionais, em que as probabilidades para cada categoria resposta foram:

π_{hijkm1} - Probabilidade de nenhum conidióforo de *B. cinerea* presente na folha;

π_{hijkm2} - Probabilidade de baixo número de conidióforos de *B. cinerea* presentes na folha;

π_{hijkm3} - Probabilidade de alto número de conidióforos de *B. cinerea* presentes na folha.

em que: h: hospedeiro;

i: temperatura;

j: inoculação do *B. cinerea*;

k: concentração de *C. rosea*;

m: Dias

Para o ajuste deste modelo a estratégia de análise foi combinar duas categorias de resposta, obtendo uma probabilidade acumulada tal que:

$\theta_{hijkm1} = \pi_{hijkm2}$ - Probabilidade de baixo número de conidióforos de *B. cinerea* presentes na folha;

$\theta_{hijkm2} = \pi_{hijkm2} + \pi_{hijkm3}$ - Probabilidade de se encontrar conidióforos de *B. cinerea* presentes na folha, e;

$$\pi_{hijkm1} = 1 - \theta_{hijkm2}.$$

Note que para uma resposta ordinal com $L > 2$ categorias, são calculados $L-1$ logitos cumulativos, baseados nas probabilidades acumuladas.

No experimento há três categorias de resposta, e portanto, há dois logitos cumulativos:

$$\text{logit}(\theta_{hijkm1}) = \ln \frac{\pi_{hijkm2}}{\pi_{hijkm1} + \pi_{hijkm3}} \quad \text{logit}(\theta_{hijkm2}) = \ln \frac{\pi_{hijkm2} + \pi_{hijkm3}}{\pi_{hijkm1}}$$

Assim, o modelo de chances (*odds*) proporcionais que leva em consideração ambas as *odds* $\theta_{h_{ijkm1}}$ e $\theta_{h_{ijkm2}}$ simultaneamente, para cada combinação das variáveis explicativas, é:

$$\text{logit}(\theta_{h_{ijkmL}}) = \beta_{0L} + \beta'_L X_{h_{ijkm}}, \text{ em que } L \text{ indexa os dois logitos.}$$

Neste modelo o parâmetro β_{01} representa o intercepto do primeiro logito cumulativo e o parâmetro β_{02} representa o intercepto do segundo logito cumulativo. β'_L é o vetor de parâmetros dos efeitos das variáveis explicativas e $X_{h_{ijkm}}$ é o vetor representativo das variáveis explicativas.

É importante perceber que haverá $L - 1$ valores estimados para os interceptos e apenas um, e somente um, valor para cada variável explicativa do modelo, sendo este valor usado em todos os logitos. A pressuposição de chances proporcionais do modelo é que garante este único parâmetro em todos os logitos.

3.2.1 – COMANDOS DO SAS PARA O MODELO DE CHANCES PROPORCIONAIS

proc logistic data = botritis ;

Proc logistic sendo usado para o conjunto de dados chamado Botritis.

class hosp(ref="0") temp (ref="0")inoc (ref="0")conc (ref="0")/ coding = reference;

Variáveis significativas no modelo:

Hosp: Hospedeiro, sendo morango (0) a categoria de referência.

Temp: Temperatura, sendo 18°C (0) a categoria de referência.

Inoc: Inoculação do Botrytis, sendo a resposta Não (0) a categoria de referência.

Conc: Concentração de Clonostachys, sendo a concentração de 10.000 conídios (0) a categoria de referência.

model conid (descending) = hosp temp inoc conc /link=logit covb;

O modelo tem como variável resposta, em função das variáveis explicativas mencionadas acima, o número de conídios de Botrytis presentes na folha (conid) e suas respostas serão ordenadas do nível mais alto da categoria resposta(2) para o

mais baixo(0). Por isso, o uso do comando `descending`. Também faz-se o uso da função de ligação `logit`.

```
output out= log predprobs = (c i);
```

```
run;
```

3.3 – MODELO DE CHANCES PROPORCIONAIS PARCIAIS

Uma das alternativas ao ajuste do modelo quando a pressuposição de chances proporcionais é rejeitada para algumas variáveis explicativas, este modelo traz um incremento de um coeficiente (γ_i), que é o efeito de não proporcionalidade associado em cada L-ésimo logito cumulativo. Assim, cada logito é ajustado associando este efeito às variáveis não proporcionais com as demais variáveis explicativas em que a pressuposição foi satisfeita. Para análise dos dados apresentados, considere apenas a variável X_3 (Inoculação do *B. Cinerea*) como sendo proporcional. Assim, para as demais variáveis explicativas houve um incremento de um coeficiente (γ) nos dois logitos. Dessa forma, as equações que definem os dois logitos são:

$$\text{logit}(\theta_{h_{i j k m 1}}) = \text{Int}_{01} + (\text{Hp}_{11} + \gamma_{11})X_1 + (\text{Tp}_{21} + \gamma_{21})X_2 + \text{In}_3X_3 + (\text{Co}_{41} + \gamma_{41})X_4 + (\text{Di}_{51} + \gamma_{51})X_5$$

e

$$\text{logit}(\theta_{h_{i j k m 2}}) = \text{Int}_{02} + (\text{Hp}_{12} + \gamma_{12})X_1 + (\text{Tp}_{22} + \gamma_{22})X_2 + \text{In}_3X_3 + (\text{Co}_{42} + \gamma_{42})X_4 + (\text{Di}_{52} + \gamma_{52})X_5$$

3.3.1 - COMANDOS DO SAS PARA O MODELO DE CHANCES PROPORCIONAIS PARCIAIS

```
contrast ' inoc all logits ' bin1;  
contrast ' hosp on logit 1 (21 vs 0) ' bh1;  
contrast ' temp on logit 1 (21 vs 0) ' bt1;  
contrast ' conc on logit1 (21 vs 0) ' bc1;  
contrast ' hosp PO ' bh1-bh2;  
contrast ' temp PO ' bt1-bt2;  
contrast ' conc PO ' bc1-bc2;
```

Contrastes realizados com o objetivo de se verificar quais variáveis explicativas são proporcionais.

Proc NLMIXED data=botritis;

Proc NLMIXED sendo usado para o conjunto de dados chamado botritis .

parms Int2 = **1**, Int1 = **2**;

Criação dos parâmetros: interceptos 1 e 2. Atribui-se à categoria resposta 0 como sendo a referência.

h=(hosp=**1**); t=(temp=**1**); in=(inoc=**1**); c=(conc=**1**);

Criação das variáveis dummies sendo o nível zero como referência.

cp2= **1**/(**1** + exp(-(Int2 + bh2*h + bt2*t + bin1*a + bc2*c)));

Probabilidades acumuladas (LOGITO 2) seguindo o modelo para intercepto 2.

cp1= **1**/(**1** + exp(-(Int1 + bh1*h + bt1*t + bin1*a + bc1*c)));

Probabilidades acumuladas (LOGITO 1) seguindo o modelo para intercepto 1.

Note que um único coeficiente para a variável inoculação é usado para ambos os logitos pois foi a única variável proporcional.

if conid=**2** then ip = cp2;
else if conid=**1** then ip = cp1-cp2;
else ip = **1**-cp1;

Cálculo das probabilidades individuais.

p = (ip>**0** and ip<=**1**)*ip + (ip<=**0**)***1e-8** + (ip>**1**);

Garante que os valores das probabilidades preditas são válidos.

loglik = log(p);

Definem o log da verossimilhança para ser maximizado pelo procedimento.

model conid ~ general(loglik);

id cp1-cp2;
predict ip out= cppnlmix;

Probabilidades cumulativas e probabilidades preditas das respostas observadas são salvas no arquivo saida.

run;

3.4 – MODELO DE CHANCES NÃO PROPORCIONAIS

Outra alternativa à rejeição do pressuposto de chances proporcionais, é ajustar o modelo de chances não proporcionais. Enquanto no modelo de chances proporcionais estimam-se vários interceptos e um único parâmetro que se admite válido em todos logitos, para cada variável independente, neste modelo são estimados vários interceptos bem como vários parâmetros, para cada variável explicativa em cada logito. Assim, neste modelo foram estimados múltiplos parâmetros para o intercepto e para as variáveis explicativas nos 2 logitos (K -1, onde K =3 categorias de resposta).

No logito 1, de equação

$$\text{Logito 1} = \text{Inter}_{01} + (\text{Hosp}_{11}).X_1 + (\text{Temp}_{21}).X_2 + (\text{Inoc}_{31}).X_3 + (\text{Conc}_{41}).X_4 + (\text{Dias}_{51}).X_5,$$

comparou-se: Há um número de conídios do *B.cinerea* inferior a 50 *versus* não há.

No logito 2, de equação

$$\text{Logito 2} = \text{Inter}_{02} + (\text{Hosp}_{12}).X_1 + (\text{Temp}_{22}).X_2 + (\text{Inoc}_{32}).X_3 + (\text{Conc}_{42}).X_4 + (\text{Dias}_{52}).X_5,$$

comparou-se: Há um número de conídios do *B.cinerea* superior a 50 *versus* não há.

3.4.1 - COMANDOS DO SAS PARA O MODELO DE CHANCES NÃO PROPORCIONAIS.

```
contrast 'hosp on logit1 (21 vs 0)' bh11;  
contrast 'temp on logit1 (21 vs 0)' bt11;  
contrast 'inoc on logit1 (21 vs 0)' bin11;  
contrast 'conc on logit1 (21 vs 0)' bc11;
```

```
contrast 'hosp PO'  
bh11-bh12;
```

```
contrast 'temp PO'  
bt11-bt12;
```

```
contrast 'aplic PO'  
bin11-bin12;
```

```
contrast 'conc PO'  
bc11-bc12;
```

```
contrast 'overall PO'  
bh11-bh12, bt11-bt12, bin11-bin12, bc11-bc12;
```

Contrastes realizados com o objetivo de se verificar quais variáveis explicativas são proporcionais.

```
proc nlmixed data=botritis;
```

Proc NLMIXED sendo usado para o conjunto de dados chamado botritis.

```
parms Int2 = 1, Int1 = 2;
```

Criação dos parâmetros: interceptos 1 e 2. Atribui-se à categoria resposta 0 como sendo a referência.

```
h=(hosp=1); t=(temp=1); in=(inoc=1); c=(conc=1)
```

Criação das variáveis dummies sendo o nível zero como referência.

```
cp2= 1/(1 + exp(-(Int2 + bh12*h + bt12*t + bin12*in + bc12*c )));
```

Probabilidade acumulada (LOGITO 2) seguindo o modelo para intercepto 2.

```
cp1= 1/(1 + exp(-(Int1 + bh11*c + bt11*t + bin11*in + bc11*c )));
```

Probabilidade acumulada (LOGITO 1) seguindo o modelo para intercepto 1.

Note que o fato do modelo ser o de chances não proporcionais garante dois coeficientes distintos para cada variável nos dois logitos.

```
if conid=2 then ip = cp2;  
else if conid=1 then ip = cp1-cp2  
else ip = 1-cp1;
```

Cálculo das probabilidades individuais.

```
p = (ip>0 and ip<=1)*ip + (ip<=0)*1e-8 + (ip>1);
```

Garante que os valores das probabilidades preditas são válidos.

```
loglik = log(p);
```

Definem o log da verossimilhança para ser maximizado pelo procedimento.

```
model conid ~ general(loglik);
```

```
id cp1-cp2  
predict ip out= npnlmixed
```

Probabilidades cumulativas e probabilidades preditas das respostas observadas são salvas no arquivo saída.

```
run;
```

3.5 – ESTIMAÇÃO DOS PARÂMETROS E ADEQUAÇÃO DO MODELO

A estimação dos parâmetros foi realizada pelo Método da Máxima Verossimilhança e a avaliação da adequação do modelo ajustado, por meio do Teste da Razão de Verossimilhança.

A construção da função de verossimilhança (Hosmer e Lemeshow, 2002) é ilustrada do seguinte modo:

Seja Y a variável dependente assumindo três categorias ditas Y_1 , Y_2 , e Y_3 . Com o auxílio de três variáveis binárias, atribui-se à categoria em que o valor pertence, o número 1 e às demais, o número 0.

Assim, se o valor observado pertence à segunda categoria, então $Y_1=0$, $Y_2=1$ e $Y_3=0$.

Logo, para uma amostra de n observações independentes, a função de verossimilhança é:

$$L(\beta) = \prod_{i=1}^n \pi_1(x_i)^{y_1} \pi_2(x_i)^{y_2} \pi_3(x_i)^{y_3}.$$

Aplicando-se o logaritmo natural, tem-se

$$l(\beta) = \sum_{i=1}^n y_1 [\ln \pi_1(x_i)] + y_2 [\ln \pi_2(x_i)] + y_3 [\ln \pi_3(x_i)].$$

As equações de verossimilhança são obtidas tomando-se as derivadas parciais de $l(\beta)$ em relação a cada um dos parâmetros desconhecidos e igualando-as a zero.

$$\frac{\partial l(\beta)}{\partial \beta_{jk}} = \sum_{i=1}^n x_{ki} (y_{ji} - \pi_{ji}).$$

As equações acima são não lineares nos parâmetros e requerem o emprego de processos iterativos como Newton-Raphson e escore de Fisher na sua solução.

Obtido as estimativas dos parâmetros β_j ($j = 0,1,2,\dots,p$) fez-se necessário avaliar a adequação do modelo ajustado comparando os valores observados da variável resposta com os valores preditos pelo modelo com, e sem, a covariável sob investigação. Essa comparação foi feita utilizando o Teste da Razão de Verossimilhança (TRV) expresso por:

$$\text{TRV} = -2 \log \left[\frac{L_{SC}}{L_{CC}} \right] = -2 [\log L_{SC} - \log L_{CC}] \quad (36)$$

e sob a hipótese nula de que os p coeficientes associados ao modelo são iguais a zero com distribuição Qui-quadrado e p graus de liberdade, pôde-se avaliar o quanto da variação dos dados foi explicada pela inclusão de termos no modelo a partir das diferenças de *Deviance*.

4. RESULTADOS

Ao testar os efeitos das variáveis explicativas, a variável *Dias*, que representa o número de dias em que a variável resposta foi observada após a inoculação de *B. cinerea*, foi não significativo ($p = 0,4104$). Portanto, a variável foi retirada do modelo. As demais variáveis foram significativas segundo o teste Wald e portanto, mantidas no modelo.

4.1. MODELO DE CHANCES PROPORCIONAIS

Uma vez selecionada as variáveis explicativas do modelo, o teste escore foi realizado com a finalidade de avaliar a pressuposição de chances proporcionais para ajuste do modelo. Na Tabela 2 estão apresentados os resultados para este modelo. A estatística deste teste foi de 164,83 com um valor- $p < 0,0001$ e 4 graus de liberdade, indicando, assim, a rejeição da hipótese de proporcionalidade. Desta forma, alternativas ao ajuste são os modelos de chances proporcionais parciais ou de chances não proporcionais.

Tabela 2 - Estimativas do modelo de chances proporcionais do experimento no qual foi avaliado o potencial de *Clonostachys rosea*, no biocontrole do mofo cinzento, causado por *Botrytis cinerea*, nas culturas de morangueiro e tomateiro. Viçosa, 2010

Parâmetro	GL	Estimação	Desvio Padrão	Wald (χ^2)	Valor-p
Intercepto 1	1	4,4469	0,3184	195,0546	<0,0001
Intercepto 2	1	0,3560	0,2124	2,8089	0,0937
X ₁ (hosp)	1	-1,1367	0,2000	32,2903	<0,0001
X ₂ (temp)	1	-2,1474	0,2172	97,7765	<0,0001
X ₃ (inoc)	1	0,6571	0,1947	11,3884	0,0007
X ₄ (conc)	1	-2,4924	0,2313	116,0795	<0,0001
Teste Escore ^{1*}	4	-	-	164,8321	<0,0001

1* Teste escore para a hipótese de chances proporcionais, depois de excluída a variável dias do modelo.

4.2. MODELO DE CHANCES NÃO PROPORCIONAIS

Após a rejeição da hipótese de proporcionalidade, os parâmetros foram estimados para o modelo de chances não proporcionais. Neste modelo foram estimados múltiplos parâmetros para o intercepto, bem como para as covariáveis nos dois logitos (k -1, onde k =3 categorias de resposta). No logito 1, de equação:

$$\text{Logito 1} = \text{Inter}_{01} + (\text{Hosp}_{11}) X_1 + (\text{Temp}_{21}) X_2 + (\text{Inoc}_{31}) X_3 + (\text{Conc}_{41}) X_4$$

representa: Há um número de conídios do *B.cinerea* inferior a 50 *versus* não há.

No logito 2, de equação:

$$\text{Logito 2} = \text{Inter}_{02} + (\text{Hosp}_{12}).X_1 + (\text{Temp}_{22}).X_2 + (\text{Inoc}_{32}).X_3 + (\text{Conc}_{42}).X_4$$

representa: Há um número de conídios do *B.cinerea* superior a 50 *versus* não há.

Uma nova análise dos efeitos das variáveis explicativas, com a exclusão da variável *Dias* do modelo, foi feita sob os cuidados de um teste de significância das mesmas, podendo ser retiradas do modelo aquelas que não derem significativas. Todas as variáveis se mostraram importantes ($p < 0,01$).

Portanto, fez-se necessário o ajuste do modelo de chances não proporcionais sob os cuidados do teste da razão de verossimilhança (544,8 ; $< 0,001$) bem como a estimação dos coeficientes e das razões de chances tendo a variável resposta $Y = 0$ como referência.

Tabela 3 - Estimativas do modelo de chances não proporcionais do experimento no qual foi avaliado o potencial do *Clonostachys rosea*, no biocontrole do mofo cinzento, causado por *Botrytis cinerea*, nas culturas de morangueiro e tomateiro. Viçosa, 2010.

Parâmetros	GL	Estimativas	Desvio Padrão	Wald (χ^2)	Valor-p	OR
Intercepto 1	512	3,7968	0,3810	9,97	<0.0001	
X ₁₁ (hosp)	512	-1,3351	0,1627	-8,20	<0.0001	0,26
X ₂₁ (temp)	512	-2,1997	0,2967	-7,41	<0.0001	0,11
X ₃₁ (inoc)	512	1,0495	0,2691	3,90	0.0001	2,85
X ₄₁ (conc)	512	-1,3351	0,1627	-8,20	<0.0001	0,26
Intercepto 2	512	2,8939	0,6084	4,76	<0.0001	
X ₁₂ (hosp)	512	-7,3505	1,1601	-6,34	<0.0001	0,00
X ₂₂ (temp)	512	-3,9976	0,6423	-6,22	<0.0001	0,02
X ₃₂ (inoc)	512	0,4926	0,3669	1,34	0.1800	1,64
X ₄₂ (conc)	512	-4,1278	0,6442	-6,41	<0.0001	0,02

Assim, o modelo de chances não proporcionais ajustado para os dois logitos é:

$$\text{Logito 1} = 3,7968 - 1,3351.\text{Hosp}_{11} - 2,1997.\text{Temp}_{21} + 1,0495.\text{Inoc}_{31} - 1,3351.\text{Conc}_{41}$$

$$\text{Logito 2} = 2,8939 - 7,3505.\text{Hosp}_{12} - 3,9976.\text{Temp}_{22} + 0,4926.\text{Inoc}_{32} - 4,1278.\text{Conc}_{41}$$

Algumas razões de chances e interpretações relevantes ao pesquisador:

Comparação do morangueiro *versus* tomateiro no logito 1

$$\frac{\text{Exp}(3,7968)}{\text{Exp}(3,7968 - 1,3351)} = \text{Exp}(1,3351) = \mathbf{3,80} \text{ ou}$$

$$\frac{\text{Exp}(3,7968 - 2,1997)}{\text{Exp}(3,7968 - 1,3351 - 2,1997)} = \text{Exp}(1,3351) = \mathbf{3,80} \text{ ou}$$

$$\frac{\text{Exp}(3,7968 - 2,1997 + 1,0495)}{\text{Exp}(3,7968 - 1,3351 - 2,1997 + 1,0495)} = \text{Exp}(1,3351) = \mathbf{3,80}$$

Interpretação: Em morangos, a chance de se encontrar um número de conídio de *B. cinerea* inferior a 50 é 3,80 vezes a de se encontrar em tomates, ou seja, os tomateiros são mais receptivos ao manejo da doença pelo Cr que os morangueiros.

Comparação da temperatura 18°C *versus* 26°C no logito 1

$$\frac{\text{Exp}(3,7968)}{\text{Exp}(3,7968 - 2,1997)} = \text{Exp}(2,1997) = \mathbf{9,02} \text{ ou}$$

$$\frac{\text{Exp}(3,7968 - 1,3351)}{\text{Exp}(3,7968 - 1,3351 - 2,1997)} = \text{Exp}(2,1997) = \mathbf{9,02} \text{ ou}$$

$$\frac{\text{Exp}(3,7968 - 1,3351 + 1,0495 - 1,3351)}{\text{Exp}(3,7968 - 1,3351 - 2,1997 + 1,0495 - 1,3351)} = \text{Exp}(2,1997) = \mathbf{9,02}$$

Interpretação: À 18°C, a chance de se encontrar um número de conídio de *B. cinerea* inferior a 50 é 9,02 vezes a de se encontrar a 26°C, ou seja, em temperaturas mais baixas, o mofo cinzento se torna mais severo.

Comparação da inoculação de *B. cinerea* versus não inoculação do *B. cinerea* no logito 1

$$\frac{\text{Exp}(3,7968 - 1,3351 - 2,1997 + 1,0495 - 1,3351)}{\text{Exp}(3,7968 - 1,3351 - 2,1997 - 1,3351)} = \text{Exp}(1,0495) = \mathbf{2,85} \text{ ou}$$

$$\frac{\text{Exp}(3,7968 + 1,0495)}{\text{Exp}(3,7968)} = \text{Exp}(1,0495) = \mathbf{2,85} \text{ ou}$$

$$\frac{\text{Exp}(3,7968 - 1,3351 - 2,1997 + 1,0495)}{\text{Exp}(3,7968 - 1,3351 - 2,1997)} = \text{Exp}(1,0495) = \mathbf{2,85}$$

Interpretação: Quando submetido à inoculação do *B. Cinerea*, a chance de se encontrar um número de conídio de *B. cinerea* inferior a 50 é 2,85 vezes a de encontrar quando não inoculado, ou seja, se torna mais difícil para o *C. rosea* controlar o doença quando há uma dosagem maior do *B. cinerea*. O fato de encontrar o *B. cinerea* mesmo quando não houve inoculação pode ser explicado por uma contaminação durante o experimento.

Comparação da concentração de 10.000 versus concentração de 10.000.000 (*C. rosea*) no logito 1

$$\frac{\text{Exp}(3,7968)}{\text{Exp}(3,7968 - 1,3351)} = \text{Exp}(1,3351) = \mathbf{3,80} \text{ ou}$$

$$\frac{\text{Exp}(3,7968 + 1,0495)}{\text{Exp}(3,7968 + 1,0495 - 1,3351)} = \text{Exp}(1,3351) = \mathbf{3,80} \text{ ou}$$

$$\frac{\text{Exp}(3,7968 - 1,3351 - 2,1997 + 1,0495)}{\text{Exp}(23,7968 - 1,3351 - 2,1997 + 1,0495 - 1,3351)} = \text{Exp}(1,3351) = \mathbf{3,80}$$

Interpretação: Quando houve aplicação de 10.000 conídios de *C. rosea*, a chance de se encontrar um número de conídio de *B. cinerea* inferior a 50 é 3,80 vezes a de encontrar quando submetido à aplicação de 10.000.000 conídios, ou seja, à medida que se aumenta a concentração de conídios de *C. rosea*, aumenta-se a eficiência do controle do mofo cinzento.

Comparação do morangueiro versus tomateiro no logito 2

$$\frac{\text{Exp}(2,8939)}{\text{Exp}(2,8939-7,3505)} = \text{Exp}(7,3505) = \mathbf{1556,97}$$

Interpretação: Em morangos, a chance de se encontrar um número superior a 50 conídios do *B. cinerea* é 1556,97 vezes a de se encontrar em tomates, ou seja, *C. rosea* é muito mais eficiente em reduzir a colonização de *B. cinerea* em folhas de tomateiro.

Uma outra alternativa à rejeição da pressuposição de chances proporcionais é o ajustamento do modelo de chances proporcionais parciais.

4.3. MODELO DE CHANCES PROPORCIONAIS PARCIAIS

Neste modelo, procurou-se investigar quais variáveis explicativas não são proporcionais (Tabela 4).

Tabela 4 - Efeitos dos contrastes realizados para investigação da pressuposição de chances proporcionais das variáveis explicativas.

Contrastes	Valor F	Valor-p
hosp on logit1 (21 vs 0)	67,31	< 0,0001
temp on logit1 (21 vs 0)	54,95	< 0,0001
inoc on logit1 (21 vs 0)	15,22	< 0,0001
conc on logit1 (21 vs 0)	67,31	< 0,0001
hosp PO	26,39	< 0,0001
temp PO	6,55	< 0,0001
inoc PO	1,69	0,1942
conc PO	17,85	< 0,0001
Overall PO	9,04	< 0,0001

*Onde se lê (21 vs 0) entenda que a categoria de resposta usada como referência foi a categoria “zero” sobre os logitos 1 e 2. PO significa *proportional odds*.

Dessa forma, apenas a variável X_3 (*Inoculação do B. cinerea*) se mostrou proporcional e tendo portanto, um único parâmetro para os dois logitos. Posteriormente, faz-se necessário um novo teste de proporcionalidade para as variáveis do modelo já com a variável X_3 (*Inoculação do B. cinerea*) como proporcional (Tabela 5).

Tabela 5 - Efeitos dos contrastes realizados para investigação da pressuposição de chances proporcionais das variáveis explicativas já com a variável Inoculação do *B. cinerea* como uma variável proporcional.

Contrastes	Valor F	Valor-p
hosp on logit1 (21 vs 0)	14,36	< 0,0001
temp on logit1 (21 vs 0)	53,88	< 0,0001
inoc all logits	15,06	< 0,0001
conc on logit1 (21 vs 0)	65,82	< 0,0001
hosp PO	50,34	< 0,0001
temp PO	7,80	< 0,0001
conc PO	5,19	< 0,0001

Assim, para as demais variáveis significativas X_1 , X_2 e X_4 , houve um incremento de um coeficiente (γ) nos dois logits e foram estimados dois parâmetros, um para cada logito.

Logo, as equações para os dois logits são:

$$\text{logit}(\theta_{hijk1}) = \text{Inter}_{01} + (\text{Hosp}_{11}).X_1 + (\text{Temp}_{21}).X_2 + (\text{Inoc}).X_3 + (\text{Conc}_{41}).X_4$$

e

$$\text{logit}(\theta_{hijk2}) = \text{Inter}_{02} + (\text{Hosp}_{12}).X_1 + (\text{Temp}_{22}).X_2 + (\text{Inoc}).X_3 + (\text{Conc}_{42}).X_4$$

em que no logito 1 fez-se a comparação: Há um número inferior a 50 conídios de *B. cinerea* presentes na folha versus não há. No logito 2 a comparação foi: Há um número qualquer de conídios de *B. cinerea* presentes na folha versus não há.

Na Tabela 6 estão apresentados os parâmetros estimados para o modelo ajustado. (TRV = 531,4).

Tabela 6 - Parâmetros do modelo de chances proporcionais parciais do experimento no qual foi avaliado o potencial do *Clonostachys rosea*, no biocontrole do mofo cinzento, causado por *Botrytis cinerea*, nas culturas de morangueiro e tomateiro. Viçosa, 2010

Parâmetros	GL	Estimativas	Desvio Padrão	Valor t	Valor-p
Intercepto 1	512	3,4224	0,3998	8,56	<0,0001
X ₁₁ (hosp)	512	1,0204	0,2692	3,79	<0,0001
X ₂₁ (temp)	512	-2,2333	0,3042	-7,34	<0,0001
X ₄₁ (conc)	512	-2,7081	0,3338	-8,11	<0,0001
Intercepto 2	512	2,7510	0,5966	4,61	<0,0001
X ₁₂ (hosp)	512	-7,4737	1,1647	-6,42	<0,0001
X ₂₂ (temp)	512	-4,1875	0,6444	-6,50	<0,0001
X ₄₂ (conc)	512	-4,3378	0,6470	-6,70	<0,0001
X ₃ (inoc ^{1/})	512	0,9110	0,2347	3,88	<0,0001

^{1/} Variável preditora proporcional, parâmetro único para os dois logitos.

Assim, o modelo de chances proporcionais parciais ajustado para os dois logitos é:

$$\text{Logito 1} = 3,4224 + 1,0204.\text{Hosp}_{11} - 2,2333.\text{Temp}_{21} + 0,9110.\text{Inoc}_{31} - 2,7081.\text{Conc}_{41}$$

$$\text{Logito 2} = 2,7510 - 7,4737.\text{Hosp}_{12} - 4,1875.\text{Temp}_{22} + 0,9110.\text{Inoc}_{32} - 4,3378.\text{Conc}_{41}$$

Algumas razão de chances e interpretações relevantes ao pesquisador:

Comparação do morangueiro *versus* tomateiro no logito 2

$$\frac{\text{Exp}(2,7510)}{\text{Exp}((2,7510 - 7,4737))} = \text{Exp}(7,4737) = \mathbf{1761,11} \text{ ou}$$

$$\frac{\text{Exp}(2,7510 - 4,1875 - 4,3378)}{\text{Exp}(2,7510 - 7,4737 - 4,1875 - 4,3379)} = \text{Exp}(7,4737) = \mathbf{1761,11}$$

Interpretação: Em morangos, a chance de se encontrar conídios de *B. cinerea* é 1761,11 vezes a de se encontrar em tomates, ou seja, os tomateiros são mais receptivos ao manejo da doença pelo Cr que os morangueiros.

Comparação da temperatura de 18°C versus 26°C no logito 1

$$\frac{\text{Exp}(3,4224)}{\text{Exp}(3,4224-2,2333)} = \text{Exp}(2,2333) = \mathbf{9,33} \text{ ou}$$

Interpretação: À 18°C, a chance de se encontrar um número inferior a 50 conídios de *B. cinerea* é 9,33 vezes a de se encontrar à 26°C, ou seja, em temperaturas mais altas, o controle da doença pelo Cr se torna mais eficiente.

Comparação da inoculação de *B. cinerea* versus não inoculação do *B. cinerea* no logito 1

$$\frac{\text{Exp}(3,4224+0,9110)}{\text{Exp}(3,4224)} = \text{Exp}(0,9110) = \mathbf{2,47}$$

Interpretação: Quando submetido à inoculação do *B. cinerea*, a chance de se encontrar um número inferior a 50 conídios de *B. cinerea* é 2,47 vezes a de se encontrar quando não inoculado, ou seja, foi mais difícil para o *C. rosea* em controlar a doença quando há uma dosagem maior do *B. cinerea*.

Comparação da concentração de 10.000 versus concentração de 10.000.000 (*C. rosea*) no logito 2

$$\frac{\text{Exp}(2,7510)}{\text{Exp}(2,7510-4,3378)} = \text{Exp}(4,3378) = \mathbf{76,54}$$

Interpretação: Quando houve aplicação de 10.000 conídios de *C. rosea*, a chance de se encontrar conídios de *B. cinerea* é 76,54 vezes a de se encontrar a uma concentração de 10.000.000 conídios, ou seja, à medida que se aumenta a concentração de conídios de *C. rosea*, aumenta-se a eficiência do controle do mofo cinzento.

Comparação do morangueiro versus tomateiro no logito 2

$$\frac{\text{Exp}(2,7510)}{\text{Exp}(2,7510-7,4737)} = \text{Exp}(7,4737) = \mathbf{1761,11}$$

Interpretação: Em morangos, a chance de se encontrar algum conídio do *B. cinerea* é 1761,11 vezes a de se encontrar em tomates.

4.4. ANÁLISE DOS MODELOS DE CHANCES PROPORCIONAIS PARCIAIS E NÃO PROPORCIONAIS

As análises entre os modelos de chances não proporcionais e proporcionais parciais estão apresentadas na Tabela 7.

Tabela 7 - Análises comparativas dos modelos de chances proporcionais parciais e não proporcionais do experimento no qual foi avaliado o potencial do *Clonostachys rosea*, no biocontrole do mofo cinzento, causado por *Botrytis cinerea*, nas culturas de morangueiro e tomateiro. Viçosa, 2010

Parâmetros	Modelos de chances proporcionais parciais	Modelos de chances não proporcionais
Intercepto 1	3,4224	3,7968
Intercepto 2	2,7510	2,8939
Temp 11	-2,2333	-2,1997
Temp 12	-4,1875	-3,9976
Inoc 11	0,9110	1,0495
Inoc 12	0,9110	0,4926
Conc 11	-2,7081	-1,3351
Conc 12	-4,3378	-4,1278
Hosp 11	1,0204	-1,3351
Hosp 12	-7,4737	-7,3505

Os dois modelos se mostraram capazes de prever resultados relevantes ao pesquisador. Os coeficientes apresentados foram semelhantes e apresentaram as mesmas interpretações nas razões de chances.

O ajuste do modelo de chances proporcionais parciais como alternativa à rejeição da pressuposição de chances proporcionais se mostrou relevante pois a variável Inoculação de *B. cinerea* se apresentou como proporcional.

Desta forma, a escolha do pesquisador pelo modelo de chances não proporcionais, apenas pelo fato desta rejeição do modelo, não se torna coerente sem antes testar a mesma pressuposição diante das variáveis.

5. DISCUSSÃO

Diante do exposto, a Regressão Logística Politémica Ordinal mostra sua contribuição como ferramenta estatística na predição de valores. Na análise do experimento, fica evidente que o intervalo onde o número de conidióforos do *B. cinerea* presentes na folha menos comum é aquele superior a 50. Isso mostra a eficácia de *C. rosea* no controle do mofo cinzento e condiz com Cotta(2010) ao citar que *C. rosea* tem potencial em se tornar um produto comercial a ser desenvolvido e usado no controle biológico da doença. A variável *Dias*, que representa dias após a inoculação, pode ser retirada do modelo ($p = 0,4104$) de modo que sua significância possa ser desprezada na predição da variável resposta. Na prática, isso significa dizer que o número de dias após a inoculação em que o pesquisador irá observar é independente do número de conidióforos de *B. cinerea* presentes na folha da cultura ou melhor ainda, é rápido o processo em que o *C. rosea* controla o *B. cinerea* já que se torna irrelevante ser checado a variável resposta com 2 ou 23 dias. Fato que não ocorreu com as outras variáveis explicativas incluídas no modelo.

Estabelecidas as variáveis explicativas, fez-se o uso dos modelos de chances proporcionais parciais e não proporcionais já que o teste de proporcionalidade acusou rejeição($p < 0,001$) da pressuposição de chances proporcionais, fato que já era esperado por ser difícil de se encontrar, na prática, um experimento que não seja rejeitada essa pressuposição. Segundo Hosmer e Lemeshow (2002) esse teste nos mostra que as variáveis explicativas tem efeitos diferentes nos logitos apresentados.

Com os ajustes dos modelos, as estimativas dos coeficientes bem como as razões de chance foram calculadas e interpretadas seguindo um interesse do pesquisador.

Hosmer e Lemeshow (2002) recomendam fazer uma análise univariada para seleção dos efeitos principais para que possa incluir no modelo apenas as variáveis significativas com um nível de significância já escolhido. Logo depois, testar o ajuste do modelo por meio de testes adequados e gráficos de resíduos e, por fim, interpretar o modelo por meio da estimativa das razões de chance. Porém, Abreu (2009) e outros autores dizem ser escassos os métodos para verificação do ajuste do modelo ordinal. As estatísticas de diagnóstico existentes, propostas por Harrel (2001) e aplicáveis aos modelos de chances proporcionais são apenas gráficos traçados a partir de regressões binárias separadas para o ponto de corte da variável ordinal. Essa

análise, segundo Abreu (2009), é extremamente útil para os modelos ordinais pois elas checam linearidade, indicando eventuais transformações que devem ser utilizadas, bem como o pressuposto de chances proporcionais com auxílio do teste escore.

Os modelos e suas interpretações simples trouxeram resultados semelhantes e se mostraram extremamente condizentes com a análise feita por Cota et al. (2010). Para os autores, a aplicação de *C. rosea* foi tão ou mais eficiente que a de fungicidas no controle do mofo cinzento e, que em programas de manejo da doença, deve-se priorizar maior concentração de *C.rosea*, visto os melhores resultados obtidos. Fato que é facilmente interpretado, por exemplo, pelo modelo logístico de chances proporcionais parciais no qual afirma que quando houve aplicação de 10.000 conídios de *C. rosea*, a chance de se encontrar conídios de *B. cinerea* é 76,54 vezes a de se encontrar a uma concentração de 10.000.000 conídios.

Outro resultado importante obtido por Cotta et al. (2010) e previsto pelos modelos é que o *C. rosea* se torna mais eficiente à 26°C que a 18°C.

Assim, consegue-se mostrar as ideias que fazem deste modelo ser capaz não tão somente apresentar qual combinação de variáveis explicativas é melhor quanto ter o poder de dimensionar quantitativamente sua significância na variável resposta. Por fim, os modelos ordinais, ainda que necessite de uma boa implementação computacional e o domínio dos comandos para execução, mostraram-se capazes de analisar dados categóricos e recomenda-se que sejam evitados procedimentos mais simples tais como a dicotomização da variável resposta e a desconsideração da ordenação dos dados, que tem como consequência a perda de informação proporcionada pelos mesmos.

Para trabalhos futuros, sugere-se uma comparação entre os modelos ordinais, respeitando cada um de seus pressupostos e comparando-os através de critérios como AIC de tal maneira que se possa estabelecer uma hierarquia entre os modelos.

6. CONCLUSÃO

- A aplicabilidade do modelo de regressão logística politômica ordinal se mostrou acessível e de fáceis interpretações.

- Há diferenças entre os modelos de chances proporcionais, não proporcionais e proporcionais parciais. A escolha do melhor modelo é dependente do caráter da variável ordinal, adequação do modelo às suposições, qualidade do ajuste e capacidade de boa explicação com reduzido número de parâmetros a serem estimados.

- Os ajustes dos modelos de chances não proporcionais e proporcionais parciais se mostraram boas alternativas à rejeição de proporcionalidade avaliada pelo teste Escore.

- As estimativas dos coeficientes dos modelos bem como das razões de chances foram interpretadas em termos práticos para a Fitopatologia mostrando o potencial do *C. rosea* em se tornar um produto comercial a ser desenvolvido e usado no controle biológico da doença, pois a aplicação de *C. rosea* foi tão ou mais eficiente do que a utilização de fungicidas no controle do mofo cinzento.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- ABREU, M.N.S.; SIQUEIRA, A.L.; CAIAFFA, W.T. Regressão Logística Ordinal em estudos epidemiológicos. **Rev. Saude Publica**, v.43, n.1, p.183-194, 2009.
- AGRESTI, A. **Categorical data Analysis**. New York: John Wiley & Sons, 1990.
- ALLISON, P.D. **Logistic Regression using the SAS System: Theory and Application**. Cary, NC: SAS Institute Inc., 1999.
- ANDERSON, J.A. Regression and Ordered Categorical Variables. **Journal of the Royal Statistical Society**, Series B, v.46, p.1-30, 1984.
- CORRAR, L.J.; PAULO, E.; FILHO, J.M.D. **Análise Multivariada**. São Paulo: Atlas, 2007.
- COTA, L.V. **Clonostachys rosea no controle biológico e manejo integrado do mofo cinzento do morangueiro**. 2010. 85f. Tese (Doutorado em Fitopatologia) – Universidade Federal de Viçosa, Viçosa.
- COX, D.R.; SNELL, E.J. **The Analysis of Binary Data**. Second Edition. London: Chapman and Hall, 1989.
- COX, D.R. **The analysis of Binary Data**. Methuen, London, 1970.
- DAY, N.; KERRIDGE, D. A general maximum likelihood discriminant. **Biometrics**, v. 23, p. 313-324, 1967.
- FENTERSTOCK, F. **Credit Scoring and the next step**. Business Credit, v.107. New York, 2005.
- GIOLO, S.R. **Análise de Dados Discretos**. Curitiba: Universidade Federal do Paraná, 2004.
- GOURIEROUX E MONFORT, **Statistics and Econometric Models**, Cambridge University Press, 1995.
- HARREL, F.E.Jr. **Regression Modelling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis**. Springer-Verlag, New York, 2001.
- HENDRICKX J. Special restrictions in multinomial logistic regression. **Stata Technical Bulletin**, v.56, p.18-26, 2000.
- HOSMER, D.W.; LEMESHOW, S. **Applied Logistic Regression**. New York: John Wiley & Sons, 2002.
- JOHNSON, R. A.; WICHERN, D.W. **Applied Multivariate Statistical Analysis**. 6th ed. New Jersey: Prentice Hall inc., 2007.
- LIMA, J.D. **A análise econômico-financeira de empresas sob a ótica da estatística multivariada**. 2002. 192f. Dissertação (Mestrado em Métodos Numéricos em Engenharia) - Universidade Federal do Paraná, Curitiba.

- McCULLAGH P. Regression models for ordinal data. Chicago: **Journal of the Royal Statistical Society**: series B, v.42, p. 109-142, 1980.
- McLACHLAN, G. **Discriminant Analysis and Statistical Pattern Recognition**. New York: John Wiley & Sons. 1992.
- MONTGOMERY, D.C.; RUNGER, G.C. **Estatística aplicada e probabilidade para engenheiros**. Rio de Janeiro: LTC, 2003.
- PAULA, G.A. **Modelos de Regressão com Apoio Computacional**. 2002.
- SAS INSTITUTE INC. **SAS/STAT® 9.2 - User's Guide**. Cary, NC. SAS Institute Inc, 2008.
- TRUETT, J.; CORNFIELD, J.; KANNEL, W. A multivariate analysis of the risk of coronary heart disease in Framingham. **Journal of Chronic Diseases**, v. 20, p. 511-524, 1967.
- WALD, T. Tests of statistical hypotheses concerning several parameters when the number of observation is large. **Trans. Amer. Math. Soc.**, v. 54, p. 426-482, 1943.
- WALKER, S. H.; DUNCAN, D. B. Estimation of the probability of an event as a function of several independent variables. **Biometrika**, v. 54, p. 167-169, 1967.
- WEISBERG, S. **Applied Linear Regression**, 3th ed. New York: Wiley. Web reference: www.stat.umn.edu/alr, 2005.