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

**Modelos de transição com dados categorizados, em processos de
tempo discreto e contínuo, com aplicações à entomologia**

Laura Vicuña Torres de Paula

Tese apresentada para obtenção do título de Doutora
em Ciências. Área de concentração: Estatística e Ex-
perimentação Agronômica

**Piracicaba
2021**

Laura Vicuña Torres de Paula
Bacharela em Estatística

**Modelos de transição com dados categorizados, em processos de
tempo discreto e contínuo, com aplicações à entomologia**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Prof. Dr. **IDEMAURO ANTONIO RODRIGUES
DE LARA**

Tese apresentada para obtenção do título de Doutora
em Ciências. Área de concentração: Estatística e Ex-
perimentação Agronômica

Piracicaba
2021

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

Paula, Laura Vicuña Torres de

Modelos de transição com dados categorizados, em processos de tempo discreto e contínuo, com aplicações à entomologia / Laura Vicuña Torres de Paula. -- versão revisada de acordo com a resolução CoPGr 6018 de 2011. -- Piracicaba, 2021 .

49 p.

Tese (Doutorado) -- USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Modelagem 2. Dados entomológico 3. Dados longitudinais 4. Processo estocástico. I. Título.

DEDICATÓRIA

A Deus, por sempre guiar meu caminho da melhor forma possível.

À minha mãe Engrácia, por seu apoio incondicional e incentivo diário.

Aos meus irmãos, Emanuel, Emanuela, Vivianny e Luiz Antônio.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por sempre iluminar meu caminho com saúde, sabedoria e determinação. E também por introduzir pessoas especiais em minha vida, que apoiam e guiam minhas escolhas.

À minha mãe Engrácia, por todas as ligações diárias com sua alegria encantadora que estarão sempre gravadas em meu coração.

A meu orientador Idemauro Antonio Rodrigues de Lara, por sua orientação e paciência na elaboração deste trabalho. Sou muito grata pelo apoio e confiança.

À professora Carolina Reigada Montoya e ao colega Vitor José Bon, por auxiliar e fornecer os insetos e ovos utilizados para desenvolver o experimento deste trabalho.

Aos meus irmãos, Emanuel, Emanuela, Vivianny e Luiz Antônio, por serem meus amigos e tudo mais que uma pessoa possa ser à outra.

Aos meus lindos sobrinhos, Ana Luzia, José Guilherme e Maria Tereza, por me terem em seus corações e por me permitirem fazer parte de suas vidas.

A todos os professores do programa de pós-graduação em Estatística e Experimentação Agronômica do Departamento de Ciências Exatas, pelos ensinamentos profissionais e pessoais.

Aos funcionários do Departamento de Ciências Exatas, que sempre tiveram muita paciência e simpatia e por sempre me ajudarem, quando ao seu alcance.

À minha grande amiga Janaina, e à mais nova amiga Tatiana que estiveram juntas comigo nessa caminhada, sempre me ajudando e incentivando.

Às minhas amigas, Kelly, Raphaela, Raquel e Wanessa que, mesmo na distância, apoiaram -me, incentivando-me e dando-me força aos meus sonhos e objetivos.

À CAPES, pelo apoio financeiro.

EPÍGRAFE

“A educação é a arma mais poderosa que você pode usar para mudar o mundo”.

Nelson Mandela

“A educação tem raízes amargas, mas os seus frutos são doces”

Aristóteles

SUMÁRIO

Resumo	7
Abstract	8
1 Introdução	9
Referências	10
2 Teste gradiente para avaliar a homogeneidade das probabilidades de transição em modelos de tempo discreto com aplicação à entomologia	11
Resumo	11
2.1 Introdução	11
2.2 Processo estocástico de tempo discreto	12
2.2.1 Modelo de transição de tempo discreto	13
2.2.2 Modelo de transição dos logitos generalizados	15
2.2.3 Modelo de transição de chances proporcionais	16
2.2.4 Teste de estacionariedade para probabilidades de transição	16
2.2.4.1 Estatística gradiente	18
2.3 Estudo de simulação	19
2.3.1 Dados categóricos nominais	20
2.3.2 Dados categóricos ordinais	21
2.4 Estudo na Entomologia	22
2.4.1 Resultados	23
2.5 Conclusão	24
Referências	25
3 Modelos de multiestados no estudo comportamental do ácaro predador <i>Phytoseiulus persimilis</i> e da microvespa <i>Telenomus podisi</i> para controle biológico de pragas	27
Resumo	27
3.1 Introdução	27
3.2 Processos estocásticos e modelos multiestados	29
3.2.1 Cadeias de <i>Markov</i> de tempo contínuo	29
3.2.2 Modelo multiestado	30
3.3 Material e métodos	32
3.3.1 Material	32
3.3.1.1 Estudo sobre o comportamento do ácaro <i>Phytoseiulus persimilis</i>	32
3.3.1.2 Estudo sobre o comportamento da microvespa <i>Telenomus podisi</i>	33
3.3.2 Métodos	34
3.4 Resultados e discussões	35
3.4.1 Análise dos dados sobre o ácaro <i>Phytoseiulus persimilis</i>	35
3.4.2 Análise dos dados sobre o parasitoide <i>Telenomus podisi</i>	39
3.5 Conclusão	42
Referências	42
4 Considerações finais	45
5 Apêndices	47
Apêndices	47
5.1 Distribuição Multinomial	47
5.1.1 Logaritmo da função de verossimilhança para modelo logito generalizado de transição	47
5.1.2 Logaritmo da função de verossimilhança para modelo chances proporcionais de transição	48

RESUMO

Modelos de transição com dados categorizados, em processos de tempo discreto e contínuo, com aplicações à entomologia

Dados com processo estocástico de tempo discreto ou contínuo podem ocorrer, em várias situações experimentais, em particular, em estudos entomológicos para avaliar comportamentos de agentes naturais que combatem pragas. As respostas desses indivíduos podem ser representadas por variáveis categóricas nominais ou ordinais. Dessa forma, este trabalho descreve classes de modelos para variáveis desta natureza, tendo como motivação experimentos desenvolvidos no Laboratório de biologia de insetos da Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo, a fim de avaliar comportamentos da microvespa *Telenomus podisi* e ácaro predador *Phytoseiulus persimilis* em estudos de controle biológico. Uma característica particular desses experimentos é o fato de descreverem ensaios com medidas repetidas ao longo tempo, uma vez que cada inseto foi avaliado em mais de uma ocasião no decorrer do tempo. Assim, modelos de transição consideram a estrutura de dados longitudinais e processos estocásticos. O trabalho está estruturado em forma de artigos. O primeiro artigo traz uma revisão sobre processo estocástico de tempo discreto e modelos de transição para analisar dados categóricos, bem como propor a utilização da estatística gradiente para verificar a condição de estacionariedade desses modelos. Sob simulação, a estatística gradiente apresentou resultados satisfatórios. Com relação ao estudo de motivação de tempo discreto, a análise permitiu identificar que a microvespa *Telenomus podisi* evita a competição intraespécie em ovos de *Euschistus heros*. O segundo artigo traz uma revisão sobre o processo estocástico de tempo contínuo e apresenta-se o modelo de multiestados, a fim de analisar dados categóricos com mensuração do tempo. Nesse artigo, apresentam-se dois experimentos entomológicos. No primeiro estudo sobre ácaro predador *Phytoseiulus persimilis*, há um maior efetividade ao combate do ácaro *Tetranychus urticae* em plantas de feijão, quando utiliza-se o composto *Beauveria bassiana* x *Metarhizium robertsii* formado por fungos entomopatogênicos. No segundo estudo sobre a microvespa *Telenomus podisi*, há maior preferência por oviposição em ovos não parasitados de *Euschistus heros*, por consequência, evita a competição intraespécie.

Palavras-chave: Dados categóricos longitudinais. Controle biológico. Probabilidades de transição. Processo estocástico. Simulação de dados.

ABSTRACT

Transition models with categorized data, in discrete and continuous time processes, with entomology applications

Data with a discrete or continuous time stochastic process can occur, in several experimental situations, in particular, in entomological studies to evaluate behaviors of natural agents that fight pests. The response actions of these individuals are represented by nominal or ordinal categorical variables. Thus, this work describes methods of models for variables of this nature, motivated by experiments developed at the Laboratory of Entomology and Biological Control of Insects of the School of Agriculture “ Luiz de Queiroz ” - University of São Paulo, in order to evaluate behaviors of the parasitoid *Telenomus podisi* and predatory mite *Phytoseiulus persimilis* in biological control studies. A particular feature of these experiments is the fact that they describe assay with repeated measures over time, since each insect has been evaluated on more than one occasion over time. Thus, transition models consider the structure of longitudinal data and stochastic processes. The work is structured in the form of articles. The first article reviews the stochastic process of discrete time and transition models to analyze categorical data, as well as proposing the use of gradient statistics to verify the stationary condition of these models. Under simulation, the gradient statistic showed satisfactory results. Regarding the discrete time motivation study, the analysis allowed to identify that the parasitoid *T. podisi* prevents intraspecies competition in *Euschistus heros* eggs. The second article reviews the stochastic process of continuous time and presents the multi-state model, in order to analyze categorical data with time measurement. In this article, two entomological experiments are presented. In the first study on predatory mite *P. persimilis*, there is greater effectiveness in combating the mite *Tetranychus urticae* in bean plants, when using the compound *Beauveria bassiana* x *Metarhizium robertsii* formed by entomopathogenic fungi. In the second study on the parasitoid *T. podisi*, there is a greater preference for oviposition in non-parasitized eggs of *E. heros*, therefore, prevents intraspecies competition.

Keywords: Longitudinal categorical data. Biological control. Transition probabilities. Stochastic process. Data simulation.

1 INTRODUÇÃO

Dados categóricos são utilizados em diversas áreas e descrevem características ou comportamentos de amostras de indivíduos em experimentos e estudos observacionais. Em alguns casos, o pesquisador tem interesse em analisar os indivíduos ao longo do tempo e tal situação resulta em dados longitudinais (Diggle et al., 2002). Nesse contexto, os modelos de transição, sendo uma extensão dos modelos lineares generalizados (Molenberghs e Verbeke, 2006), são uma alternativa para analisar dados com essa estrutura, uma vez que são capazes de avaliar a dependência entre as observações e as mudanças das categorias de resposta (Lara et al., 2017). Conforme Diggle et al. (2002), os modelos de transição são baseados em processos estocásticos. Além disso, o tempo do processo pode ser definido como discreto ou contínuo. Pode-se citar, como exemplo, o trabalho desenvolvido por Lara et al. (2017), que consiste em apresentar modelos de transição com processo discreto, para avaliar lesões em suínos. Ademais, Conn e Cooch (2009) descrevem modelos de transição com processo contínuo, para aferir a evolução de conjuntivite em pássaros.

Aplicações em modelos com essas características possibilitam mensurar comportamentos das mudanças das categorias da variável resposta em processos de tempo discreto, como também as intensidades das mudanças em modelos com processo contínuo. Uma condição importante a ser verificada nesses modelos é a questão da homogeneidade das probabilidades de transição. Diante disso, Anderson e Goodman (1957) e Lara et al. (2018) desenvolveram testes, baseados em estimativas das probabilidades de transição e dos coeficientes de regressão, respectivamente. As estatísticas Anderson e Goodman (1957), razão de verossimilhança e *wald* apresentaram bom desempenho para o erro tipo I e poder do teste (Lara et al., 2018). Introduzida por Terrell (2002), a estatística gradiente é uma alternativa às estatísticas de razão de verossimilhança e *wald*, e apresentam as mesmas propriedades assintóticas (Lemonte e Ferrari, 2012). O fato de sua expressão ser mais simples e não necessitar de cálculo de matriz inversa, torna-a mais simples de se obter. Vargas et al. (2014) ajustaram modelos lineares generalizados com pequenas amostras para apresentar resultados satisfatórios da estatística gradiente, enquanto Lemonte (2016) evidencia propriedades e comparações do poder do teste. Nesse contexto, a estatística gradiente é implementada nesse trabalho para analisar homogeneidade de probabilidades de transição.

Como aplicações práticas são considerados estudos que envolvem controle biológico de pragas. Identificar e explorar os melhores inimigos naturais de uma determinada praga são objetivos potenciais em experimentos entomológicos. Nesse contexto, a título de ilustração em jardins ou pequenas fazendas, a microvespa social foi potencialmente utilizada em combates contra pragas (Prezoto et al., 2019), enquanto fungos podem ser empregados para melhorar o crescimento e capacidade endofítica em plantas de morango, (Canassa et al., 2019). Ao explorar esses inimigos naturais, comportamentos podem ser observados e consequentemente, modelos de transição podem ser empregados, para analisar as preferências dos indivíduos, como também a velocidade/intensidade das mudanças de comportamento dos insetos sob efeito de tratamentos.

Nessas circunstâncias, um objetivo central deste trabalho é contribuir com as análises de estudos entomológicos desenvolvidos no Laboratório de Entomologia e Controle Biológico de Insetos da Escola Superior de Agricultura “Luiz de Queiroz- Universidade de São Paulo. Os experimentos caracterizam processos *markovianos* de tempo discreto e contínuo, que descrevem fenômenos aleatórios variando ao longo tempo (Ross et al., 1996). Em vista disso, os objetivos específicos deste trabalho são:

- 1) Construir um teste baseado nas estimativas dos coeficientes de regressão, utilizando a estatística gradiente, a fim de verificar homogeneidade das probabilidades de transição em modelos com processo discreto.
- 2) Apresentar a metodologia dos modelos de transição com tempo contínuo e suas principais caracte-

rísticas.

A fim de desenvolver essas propostas, o presente trabalho está estruturado da seguinte forma: Capítulo 2, revisão sobre processo estocástico, modelos de transição de tempo discreto, teste de estacionariedade para probabilidades de transição, estudo de simulação e aplicação de um estudo entomológico. No Capítulo 3, descrição sobre processo estocástico de tempo contínuo, modelos multiestados e aplicações a experimentos entomológicos. No Capítulo 4 apresentam-se as considerações finais.

Referências

- Anderson, T. W. e Goodman, L. A. (1957). Statistical inference about markov chains. *The Annals of Mathematical Statistics*, pages 89–110.
- Canassa, F., Tall, S., Moral, R. A., de Lara, I. A., Delalibera Jr, I., e Meyling, N. V. (2019). Effects of bean seed treatment by the entomopathogenic fungi *metarhizium robertsii* and *beauveria bassiana* on plant growth, spider mite populations and behavior of predatory mites. *Biological Control*, 132:199–208.
- Conn, P. B. e Cooch, E. G. (2009). Multistate capture–recapture analysis under imperfect state observation: an application to disease models. *Journal of Applied Ecology*, 46(2):486–492.
- Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Heagerty, P. J., Zeger, S., et al. (2002). *Analysis of longitudinal data*. Oxford University Press.
- Lara, I., Hinde, J., de Castro, A., e da Silva, I. (2017). A proportional odds transition model for ordinal responses with an application to pig behaviour. *Journal of Applied Statistics*, 44(6):1031–1046.
- Lara, I. A. R., Hinde, J., e Taconeli, C. A. (2018). An alternative method for evaluating stationarity in transition models. *Journal of Statistical Computation and Simulation*, 87(15):2962–2980.
- Lemonte, A. (2016). *The gradient test: another likelihood-based test*. Academic Press.
- Lemonte, A. J. e Ferrari, S. L. (2012). The local power of the gradient test. *Annals of the Institute of Statistical Mathematics*, 64(2):373–381.
- Molenberghs, G. e Verbeke, G. (2006). *Models for discrete longitudinal data*. Springer Science & Business Media.
- Prezoto, F., Maciel, T. T., Detoni, M., Mayorquin, A. Z., e Barbosa, B. C. (2019). Pest control potential of social wasps in small farms and urban gardens. *Insects*, 10(7):192.
- Ross, S. M., Kelly, J. J., Sullivan, R. J., Perry, W. J., Mercer, D., Davis, R. M., Washburn, T. D., Sager, E. V., Boyce, J. B., e Bristow, V. L. (1996). *Stochastic processes*, volume 2. Wiley New York.
- Terrell, G. R. (2002). The gradient statistic. *Computing Science and Statistics*, 34(34):206–215.
- Vargas, T. M., Ferrari, S. L., e Lemonte, A. J. (2014). Improved likelihood inference in generalized linear models. *Computational Statistics & Data Analysis*, 74:110–124.

2 TESTE GRADIENTE PARA AVALIAR A HOMOGENEIDADE DAS PROBABILIDADES DE TRANSIÇÃO EM MODELOS DE TEMPO DISCRETO COM APLICAÇÃO À ENTOMOLOGIA

Resumo

Modelos de transição, de tempo discreto ou contínuo, podem ser utilizados como alternativa para analisar dados longitudinais. Sua metodologia emprega processos estocásticos para descrever dependências entre observações, mudanças dos estados ao longo do tempo e efeito de covariáveis. Nesse trabalho o enfoque está direcionado aos modelos de tempo discreto e com variável resposta categórica nominal. Como motivação, apresenta-se um experimento na área entomológica com a espécie parasitoide *Telenomus podisi*. Esse parasitoide tem grande potencial na área agrícola, principalmente, quando empregado no combate à praga *Euchistus heros*. Nesse contexto, avaliou-se o comportamento dessa microvespa sob duas condições de tratamentos (ovos não parasitados e ovos previamente parasitados). Assim os objetivos desse estudo foi inferir sobre o potencial desse parasitoide reconhecer ovos previamente parasitados, bem como verificar a condição de estacionariedade das probabilidades das mudanças de comportamento ao longo do tempo. Para analisar a estacionariedade é proposta a estatística gradiente. Os resultados do estudo de simulação indicaram um bom desempenho dessa estatística, em relação ao erro tipo I e poder do teste, quando comparadas com teste clássico, *Wald* e razão de verossimilhança. Analisando o experimento entomológico por meio do modelo de transição, as estatísticas gradiente, *Wald*, razão de verossimilhança e clássica entraram em conformidade e indicaram a utilização de um processo estacionário. As estimativas das probabilidades de transição sinalizaram que a microvespa *Telenomus podisi* evita a competição intraespécie, em consequência, as probabilidades de oviposição em ovos previamente parasitados são menores que em ovos não parasitados.

Palavras-chave: Dados categóricos longitudinais. Máxima verossimilhança. Probabilidades de transição. Processo estocástico. Simulação de dados.

2.1 Introdução

Dados categorizados são comuns em estudos experimentais, principalmente na área de ciências agrárias (Lara et al., 2018). Qualificar o comportamento de um animal, o grau de lesão ou infestação por pragas, germinação de uma determinada planta são essenciais e, com frequência, empregados. Em algumas situações, as unidades experimentais são medidas repetidas ao longo do tempo e resultam em estudos com dados longitudinais (Diggle et al., 2002). Os modelos de transição podem avaliar a dependência entre observações e mudanças dos estados ao longo do tempo (Lara et al., 2017), além de entender como as covariáveis alteram as probabilidades dessas mudanças, as quais não são possíveis nos modelos marginais e de efeitos aleatórios (Ware et al., 1988).

Quando a variável de interesse é do tipo categórica nominal, o modelo dos logitos generalizados é o mais indicado (Agresti, 2002). No entanto, para uma variável resposta do tipo ordinal, os modelos cumulativos são os mais recomendados (McCullagh, 1980). Caso a variável resposta categórica seja ordenada, os modelos cumulativos adquirem algumas vantagens, sobretudo, com o uso do modelo de chances proporcionais, pois permite maior simplicidade e parcimônia (Lara et al., 2018).

Os modelos de transição vem sendo aplicados em diversas situações. Para dados categóricos ordinais, Yu et al. (2003) utilizaram essa técnica para verificar eficácia de diferentes estratégias para o tratamento de dor lombar em um ensaio clínico. Em outra abordagem, Ganjali e Rezaee (2007) distinguiu-se a diferença entre tipos de tratamentos aplicados a pacientes com problemas de insônia. Ademais, Lara

et al. (2017) analisou a gravidade das lesões em suínos, quando submetidos a duas condições de criação: com e sem enriquecimento ambiental. Já para dados nominais, Albert (2000) utiliza desse modelo em ensaios farmacêuticos com valores ausentes não sendo ignorados. Diferente de Yang et al. (2007), que utilizou essa classe de modelos com valores ausentes sendo ignorados e não ignorados.

Quando o interesse é verificar as mudanças de dois ou mais processos em um conjunto de dados, as propriedades multivariadas são levadas em consideração. Dessa forma, Zeng e Cook (2007) descrevem o modelo de transição para dados binários longitudinais multivariados. Para dados com comportamento serial, Heagerty (2002) desenvolveu um modelo de transição marginalizado e avaliou-o em estudos de simulações e com dados sobre pacientes esquizofrênicos. Ademais, Lee e Daniels (2007) estenderam essa classe de modelos marginalizados em um ensaio clínico de câncer de colo retal.

Os modelos de transição tem como característica comparar as matrizes de probabilidades entre os grupos de tratamentos e Ware et al. (1988) classificam-nos como condicionais por levar em conta as respostas passadas e um conjunto de covariáveis. Em modelos com cadeias markovianas, uma suposição a ser analisada é a condição de estacionariedade ou homogeneidade das probabilidades de transição ao longo do tempo. Em geral, tal condição não é atendida. Na literatura, há testes para avaliar tal condição como o teste clássico de Anderson e Goodman (1957), baseado nas probabilidades de transição estimadas e os testes baseados nos coeficientes de regressão, razão de verossimilhança e *Wald*, apresentados em Lara et al. (2017, 2018).

Portanto, o objetivo deste artigo é estender a proposta dos autores Lara et al. (2018) e apresentar o teste gradiente para verificar estacionariedade nos modelos de transição. Essa estatística foi introduzida por Terrell (2002) e é uma alternativa às estatísticas de razão de verossimilhança e *Wald*. Além disso, compartilham das mesmas propriedades assintóticas de primeira ordem (Lemonte e Ferrari, 2012). A estatística gradiente vem sendo empregada em diversos trabalhos e algumas de suas propriedades tais como, comparações entre poder do teste com outras estatísticas, correção tipo *bartlett*, robustez e aplicações são descritas por Lemonte (2016).

Em um estudo com pequenas amostras a estatística gradiente mostrou-se ser uma alternativa interessante na classe de modelos lineares generalizados (Vargas et al., 2014). Entretanto, para dados com valores extremos, Ferrari e Pinheiro (2014) ajustaram um modelo de regressão e concluíram que as estatísticas gradiente, *Wald* e razão de verossimilhança obtiveram resultados inferiores a uma estatística de razão de verossimilhança ajustada.

Com objetivo de apresentar a proposta do teste gradiente para verificar estacionariedade em modelos de transição, este artigo será estruturado da seguinte forma: na seção 2, apresenta-se uma revisão sobre processo estocástico de tempo discreto, descrição sobre modelos de transição para dados nominais e ordinais, teste de hipótese para estacionariedade e estatística gradiente. Na seção 3, descreve-se um estudo de simulação para dados nominais e ordinais. Na seção 4, apresenta-se um estudo de motivação em Entomologia, análise e discussão dos resultados. Por fim, seção 5, as considerações finais.

2.2 Processo estocástico de tempo discreto

Um processo estocástico é uma coleção de variáveis aleatórias indexadas no tempo (Ross et al., 1996) com características imprevisíveis, uma vez que podem produzir resultados diferentes em cada execução. Além disso, Karlin e Taylor (1975) descrevem um processo como uma sequência de eventos regidos por leis probabilísticas sendo de tempo discreto ou contínuo. Quando o tempo é definido como enumerável, tem-se um processo de tempo discreto, caso contrário, de tempo contínuo.

Atualmente existem vários processos (*Bernoulli*, *Markov*, *Poisson* e *Gaussiano* são exemplos) sendo o *Markov* ou *Markoviano* o mais utilizado. Sua principal característica é que, dado o estado presente, o passado não influencia no estado futuro (Hoel et al., 1986). Nesse contexto, considere um

processo estocástico Y_t de tempo discreto, com $S = \{1, 2, \dots, k\}$ representando os estados do processo e $\tau = (0, 1, \dots, T)$ o tempo, em que $t \in \tau$. Então, a probabilidade da transição do estado a para o b é dada por

$$\pi_{ab}(t-1, t) = P(Y_t = b | Y_{(t-1)} = a), \quad \forall a, b \in S.$$

sendo que essas probabilidades podem ser homogêneas ou não ao longo do tempo, caracterizando os processos estacionários e não estacionários (Lara et al., 2017). Quando as probabilidades de um indivíduo passar de um estado para outro e forem respectivamente independentes do tempo, tem-se as homogeneidades das probabilidades de transição ou estacionariedade (Lee et al., 1965). Para simplificar a notação, as probabilidades podem ser rescritas como $\pi_{ab}(t-1, t) = \pi_{ab}(t)$ e em forma matricial, tem-se

$$\mathbf{P}(t) = \begin{pmatrix} \pi_{11}(t) & \pi_{12}(t) & \dots & \pi_{1k}(t) \\ \pi_{21}(t) & \pi_{22}(t) & \dots & \pi_{2k}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{k1}(t) & \pi_{k2}(t) & \dots & \pi_{kk}(t) \end{pmatrix}$$

em que cada elemento corresponde a uma transição e satisfaz, $\pi_{ab}(t) \geq 0$, $a, b = 1, 2, \dots, k$ e $\sum_{b=1}^k \pi_{ab}(t) = 1$. Quando estacionárias ou homogêneas no tempo, denota-se por π_{ab} , ou em termos matriciais, $\mathbf{P}$, em que a ausência do parâmetro (t) indica independência em relação ao tempo.

2.2.1 Modelo de transição de tempo discreto

Considere uma amostra aleatória de N indivíduos, cada qual avaliado em n_i ocasiões pontuais, tal que $y_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$ é um vetor $(n_i \times 1)$ de variáveis respostas do i -ésimo indivíduo, $x_{it} = (x_{it1}, x_{it2}, \dots, x_{itp})^T$ é um vetor $(p \times 1)$ de covariáveis no tempo t e $h_{it} = (y_{i(t-1)}, y_{i(t-2)}, \dots, y_{i(t-q)})$, um vetor $(q \times 1)$. Conforme Diggle et al. (2002), o modelo de transição especifica um modelo linear generalizado, para uma distribuição condicionada de Y_{it} , dadas as respostas anteriores e um conjunto de covariáveis. As variáveis aleatórias $Y_{it} | \mathbf{h}_{it}$, $i = 1, \dots, N$, são condicionalmente independentes e têm distribuição pertencente à família exponencial na forma canônica

$$f(y_{it} | \mathbf{h}_{it}) = \exp \left\{ \frac{1}{\phi} [y_{it} \theta_{it} - b(\theta_{it})] + c(y_{it}, \phi) \right\} \quad (2.1)$$

em que $b(\cdot)$ e $c(\cdot)$ são funções conhecidas e que dependem de y_{it} , θ_{it} (parâmetro canônico) e ϕ (parâmetro de escala conhecido). A média e variância condicionadas são dadas por

$$\begin{aligned} E(Y_{it} | \mathbf{h}_{it}) &= b'(\theta_{it}) = \mu_{it}^C \\ \text{Var}(Y_{it} | \mathbf{h}_{it}) &= \phi b''(\theta_{it}) = v_{it}^C. \end{aligned}$$

Lara et al. (2018) sugerem que a média e variância condicionais satisfazem às seguintes equações:

$$\begin{aligned} g(\mu_{it}^C) &= \eta_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \sum_{r=1}^s f_r^*(\mathbf{h}_{it}; \boldsymbol{\alpha}) \\ v_{it}^C &= \phi(\mu_{it}^C) \end{aligned}$$

em que $g(\mu_{it}^C)$ e $v(\mu_{it}^C)$, representam respectivamente a função de ligação e de variância. Já f_r^* são funções que definem a estrutura estocástica do modelo de transição no preditor linear (Lara et al., 2018). Os parâmetros de interesse desse modelo são representados pelo vetor $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$, em que $\boldsymbol{\beta}$ tem dimensão $(p \times 1)$ e está associado às covariáveis, $\boldsymbol{\alpha}$ às respostas anteriores com dimensão definida f_r^* . De fato, segundo os autores, em alguns casos a dimensão de $\boldsymbol{\delta}$ pode ser maior que $(p+q) \times 1$. Isso acontece, quando o histórico das respostas anteriores são mais complexas ou quando existe interação entre as covariáveis (Lara et al., 2018). Entretanto, conforme Lara et al. (2017), o modelo de transição *markoviano* de primeira ordem é o mais utilizado na prática, nos quais as probabilidades de transições são preditas pelas covariáveis e pela resposta do indivíduo no tempo anterior.

Sobre o processo de estimação, Diggle et al. (2002) consideram que, em um modelo de transição com ordem q , a distribuição condicionada de $Y_{it}|\mathbf{h}_{it}$ é dada por

$$f(y_{it}|\mathbf{h}_{it}) = f(y_{it}|y_{i(t-1)}, y_{i(t-2)}, \dots, y_{i(t-q)}),$$

e a contribuição da verossimilhança do indivíduo i é dada por

$$L_i(\boldsymbol{\delta}) = f(y_{i1}, y_{i2}, \dots, y_{iq}) \prod_{j=q+1}^{n_i} f(y_{ij}|y_{i(j-1)}, y_{i(j-2)}, \dots, y_{i(j-q)}).$$

Segundo os autores, o modelo linear generalizado (2.1) representa apenas a distribuição condicional de $y_{it}|\mathbf{h}_{it}$, em que a verossimilhança das primeiras q observações não são diretamente determinadas. Consequentemente, a função verossimilhança completa não pode ser especificada. Uma alternativa para estimar $\boldsymbol{\beta}$ e $\boldsymbol{\alpha}$ seria a maximização da função de verossimilhança condicional

$$L(\boldsymbol{\delta}) = \prod_{i=1}^N f(y_{i(q+1)}, \dots, y_{in_i}|y_{i1}, \dots, y_{iq}) = \prod_{i=1}^N \prod_{t=q+1}^{n_i} f(y_{it}|\mathbf{h}_{it}).$$

Quando o processo *markoviano* é estacionário, ao longo do tempo, ajusta-se um modelo, usando contribuições individuais para a função de probabilidade, como descrito nos trabalhos Molenberghs e Verbeke (2006) e Diggle et al. (2002). Caso contrário, define-se um modelo para cada ocasião, em que $\boldsymbol{\delta}_t$ descreve a contribuição de cada transição (Ware et al., 1988), sendo baseado nos métodos de Stram et al. (1988). Adicionalmente, Lara et al. (2018) mostram que, caso $f_r^*(\mathbf{h}_{it}; \boldsymbol{\alpha}) = \alpha_r f_r^*(\mathbf{h}_{it})$, tem-se que $g(\mu_{it}^C) = \mathbf{x}_{it}^T \boldsymbol{\beta} + \sum_{r=1}^s \alpha_r f_r^*(\mathbf{h}_{it})$ é uma função linear em ambos os parâmetros $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$. Assim, é possível usar a estimação dos parâmetros do modelo linear generalizado para dados independentes. Modelos de transição com respostas multinominais são descritos como extensões multivariadas do modelo linear generalizado (Lara et al., 2017), incluindo o modelo logito generalizado para dados nominais (Agresti, 2002) e modelos de chances proporcionais para dados ordenados (McCullagh, 1980). Esses modelos

são caracterizados por incluir a cadeia de *markov* de primeira ordem (categoria de resposta observada anteriormente) como uma variável explicativa no modelo (Lara et al., 2017).

2.2.2 Modelo de transição dos logitos generalizados

A regressão logística é um modelo usual para dados categóricos dicotômicos, mas quando a variável resposta é do tipo politômica (assume mais de duas categorias de resposta) recomenda-se o modelo dos logitos generalizados (Agresti, 2002; Hosmer et al., 2013). Esse modelo tem como propriedade comparar cada categoria de resposta com uma de referência a critério do pesquisador.

Ao considerar uma variável resposta nominal assumindo valores no conjunto $S = \{1, 2, \dots, k\}$ com processo estocástico de tempo discreto, isto é, $\tau = 1, \dots, t$, a resposta para o i -ésimo indivíduo é denotada pelo vetor $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^\top$, em que as respostas do i -ésimo indivíduo representam variáveis indicadoras. Além disso, $\mathbf{x}_{it} = (x_{it1}, x_{it2}, \dots, x_{itp})^\top$ é o vetor das p variáveis explicativas associadas ao indivíduo i na t -ésima transição e $\pi_{ab}(t)(\mathbf{x})$ é a probabilidade de transição do indivíduo mudar da categoria a para b , com $\sum_{b=1}^k \pi_{ab}(t)(\mathbf{x}) = 1$. Ao fixar a k -ésima categoria como referência e considerando o modelo dos logitos generalizados (Agresti, 2002), tem-se que modelo de transição politômico nominal é dado por

$$\eta = \log \left(\frac{\pi_{ab}(t)(\mathbf{x})}{\pi_{ak}(t)(\mathbf{x})} \right) = \lambda_{b(t)} + \boldsymbol{\delta}_t^\top \mathbf{x}, \quad (2.2)$$

para um estado prévio a , em que $b = 1, \dots, k-1$, $\lambda_{b(t)}$ é o intercepto e também parâmetro de perturbação do modelo e $\boldsymbol{\delta}_t^\top = (\beta_{b1t}, \beta_{b2t}, \dots, \beta_{bpt}, \alpha_{bt})$ é o vetor de parâmetros desconhecidos a serem estimados. Conforme Agresti (2002), o intercepto $\boldsymbol{\lambda}$ e vetor $\boldsymbol{\delta}$ são divergentes para cada logito, implicando em efeitos de covariáveis diferentes para cada categoria de resposta.

As estimativas dos parâmetros desse modelo são obtidas a partir da maximização da função de verossimilhança, sendo necessário recorrer a um processo iterativo pois a equação não possui forma fechada. As probabilidades de transição são obtidas utilizando a metodologia do princípio da invariância descrita por

$$\hat{\pi}_{ab}(t)(\mathbf{x}) = \frac{\exp(\hat{\lambda}_{b(t)} + \hat{\boldsymbol{\delta}}_t^\top \mathbf{x})}{1 + \sum_{b=1}^{k-1} \exp(\hat{\lambda}_{b(t)} + \hat{\boldsymbol{\delta}}_t^\top \mathbf{x})} \quad b \in S.$$

Após obter as probabilidades estimadas, é possível montar a matriz $\hat{\mathbf{P}}(t)$, em que cada elemento da matriz corresponde a uma probabilidade de transição, descrita da seguinte forma

$$\hat{\mathbf{P}}(t) = \begin{pmatrix} \hat{\pi}_{11}(t)(\mathbf{x}) & \hat{\pi}_{12}(t)(\mathbf{x}) & \dots & \hat{\pi}_{1k}(t)(\mathbf{x}) \\ \hat{\pi}_{21}(t)(\mathbf{x}) & \hat{\pi}_{22}(t)(\mathbf{x}) & \dots & \hat{\pi}_{2k}(t)(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\pi}_{k1}(t)(\mathbf{x}) & \hat{\pi}_{k2}(t)(\mathbf{x}) & \dots & \hat{\pi}_{kk}(t)(\mathbf{x}) \end{pmatrix}. \quad (2.3)$$

Essa matriz pode ou não ser estacionária. A verificação dessa questão é realizada, por meio de testes de hipótese para estacionariedade, sendo tratado futuramente neste trabalho.

2.2.3 Modelo de transição de chances proporcionais

O modelo dos logitos cumulativos é usual para dados com respostas categóricas ordenadas, podendo ser dividido em três classes: modelos de chances não proporcionais (Williams e Grizzle, 1972), proporcionais (McCullagh, 1980) e parciais (Peterson e Harrell Jr, 1990). Quando há indícios de que os parâmetros associados às variáveis explicativas são proporcionalmente equivalentes, utiliza-se o modelo de chances proporcionais, sendo descrito neste trabalho, devido aos estudos por simulação.

Supondo uma variável resposta ordinal com tempos pré-definidos (discreto) assumindo valores no conjunto $S = \{1, 2, \dots, k\}$, em que $1 < 2 < \dots < k$, a resposta para o i -ésimo indivíduo é denotada pelo vetor $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^\top$, em que as respostas do i -ésimo indivíduo representam variáveis indicadoras. Além disso, tem-se que $\mathbf{x}_{it} = (x_{it1}, x_{it2}, \dots, x_{itp})^\top$ é vetor das p variáveis explicativas e $\pi_{ab}(t)(\mathbf{x})$ é a probabilidade de transição do indivíduo mudar da categoria a para b , com $\sum_{b=1}^k \pi_{ab}(t)(\mathbf{x}) = 1$. Assim, o modelo de transição de chances proporcionais é dado por

$$\eta = \log \left(\frac{\gamma_{ab}(t)(\mathbf{x})}{1 - \gamma_{ab}(t)(\mathbf{x})} \right) = \lambda_{b(t)} + \boldsymbol{\delta}_t^T \mathbf{x}, \quad (2.4)$$

em que $b = 1, 2, \dots, k-1$, $\lambda_{b(t)}$ é o intercepto do modelo e $\boldsymbol{\delta}_t^T = (\beta_{1t}, \beta_{2t}, \dots, \beta_{pt}, \alpha_t)$ o vetor de parâmetros desconhecidos.

As estimativas dos parâmetros desse modelo também são obtidas pela maximização da função de verossimilhança, sendo necessário recorrer a um processo iterativo. As probabilidades acumuladas e marginais de transição são obtidas utilizando o princípio da invariância descrita por

$$\hat{\gamma}_{ab}(t)(\mathbf{x}) = \frac{\exp(\hat{\lambda}_{b(t)} + \hat{\boldsymbol{\delta}}_t^T \mathbf{x})}{1 + \exp(\hat{\lambda}_{b(t)} + \hat{\boldsymbol{\delta}}_t^T \mathbf{x})},$$

em que $\hat{\gamma}_{ab}(t)(\mathbf{x}) = \hat{\pi}_{a1}(t)(\mathbf{x}) + \dots + \hat{\pi}_{ab}(t)(\mathbf{x})$.

A partir das probabilidades acumuladas, o pesquisador também consegue obter os valores das probabilidades marginais em cada categoria de resposta. Sendo definidas por

$$\hat{\pi}_{ab}(t)(\mathbf{x}) = \hat{\gamma}_{ab}(t)(\mathbf{x}) - \hat{\gamma}_{a(b-1)}(t)(\mathbf{x}) \quad a, b \in S$$

podendo serem representadas matricialmente por $\hat{\mathbf{P}}(t)$ como definido em (2.3).

2.2.4 Teste de estacionariedade para probabilidades de transição

Na análise de dados por meio dos modelos de transição para variáveis nominais e ordinais, é importante verificar se a condição de estacionariedade é satisfeita (Lai, 1967). Em caso afirmativo, há uma redução do número de parâmetros e, em consequência tem-se uma única matriz de probabilidades de transição (Lara et al., 2017). Para verificar essa condição, considere um modelo de transição com processo *markoviano*, descrito por T transições e $\pi_{ab}(t)(\mathbf{x})$ representa a probabilidade do indivíduo mover-se da categoria a para b no tempo t dado o efeito de covariáveis. As hipóteses a serem testadas são definidas da seguinte forma

$$\begin{aligned} H_0 : & \pi_{ab}(t)(\mathbf{x}) = \pi_{ab}(\mathbf{x}), \text{ para todos os } t = (1, \dots, T), \text{ e } a, b, \in (1, \dots, k); \\ H_1 : & \pi_{ab}(t)(\mathbf{x}) \neq \pi_{ab}(l)(\mathbf{x}), \text{ para algum } t \neq l, \end{aligned}$$

em que $\pi_{ab}(\mathbf{x})$ representa a probabilidade homogênea (ao longo do tempo) do indivíduo mudar da categoria a para b .

Em termos matriciais, as hipóteses são descritas da seguinte forma

$$\begin{aligned} H_0 : & \mathbf{P}(t) = \mathbf{P}, \text{ para todos os } t = (1, \dots, T); \\ H_1 : & \mathbf{P}(t) \neq \mathbf{P}(l), \text{ para algum } (t, l) \in (1, \dots, T) \text{ em que } t \neq l. \end{aligned}$$

Inicialmente, Anderson e Goodman (1957) propuseram um teste para essas hipóteses, que equivale ao teste de homogeneidade em uma tabela de contingência $(T \times k)$. Sob H_0 a estatística desenvolvida por Anderson e Goodman (1957) é descrita por

$$\xi = \sum_{t=1}^T \sum_{a=1}^k \sum_{b=1}^k \frac{n_a(t-1)[\hat{\pi}_{ab}(t)(\mathbf{x}) - \hat{\pi}_{ab}(\mathbf{x})]^2}{\hat{\pi}_{ab}(\mathbf{x})}, \quad (2.5)$$

em que $n_a(t-1)$ é o número de indivíduos que estão no estado a e tempo $(t-1)$. De acordo com os autores, para dados nominais, a estatística ξ tem distribuição assintoticamente χ_v^2 com $v = k(k-1)(T-1)$. No contexto de dados categóricos ordinais, Lara et al. (2018) salientam que é necessário definir um número adequado de graus de liberdade. Adicionalmente, para aplicar esse teste deve-se ajustar $(T+1)$ modelos de transição para, em seguida, obter as estimativas das probabilidades estimadas (Lara et al., 2018).

Como alternativas ao teste clássico de Anderson e Goodman (1957), que necessita das probabilidades estimadas, Lara et al. (2017) propuseram testes para estacionariedade baseados diretamente nos parâmetros estimados de δ_t do modelo de transição. Os autores apresentam, alternativamente, as hipóteses a serem testadas por

$$\begin{aligned} H_0 : & \delta_t = \delta_0, \text{ para todos os } t = (1, \dots, T); \\ H_1 : & \delta_t \neq \delta_l, \text{ para algum } t \neq l, \end{aligned}$$

em que δ_0 é o vetor de parâmetros do modelo de transição assumindo estacionariedade. Os testes são baseados na teoria da verossimilhança, sendo eles os teste da razão de verossimilhanças e *Wald*.

De acordo com Lara et al. (2018), ao considerar um modelo de transição com as características do modelo linear generalizado (com estrutura de dependência no preditor linear), o vetor de parâmetros a ser utilizado $\hat{\delta}_t$ tem distribuição assintoticamente normal. Assim, $\hat{\delta}_t \sim N(\delta, \mathcal{F}^{-1})$, em que $\mathcal{F}$ é a matriz de informação de *Fisher*. A estatística da razão de verossimilhança é dada por

$$\lambda = \frac{\sup_{\delta} [L(\delta, \mathbf{x} | \delta \in \Theta_0)]}{\sup_{\delta_t} [L(\delta_t, \mathbf{x} | \delta_t \in \Theta)]}, \quad (2.6)$$

em que Θ_0 representa o espaço amostral do parâmetro associado ao modelo sob H_0 , e Θ espaço amostral do parâmetro associado ao modelo de interesse. Sob as condições de regularidade e H_0 , tem-se que $\Lambda = -2\log(\lambda)$ tem distribuição assintoticamente χ_v^2 , em que $v = \dim(\Theta) - \dim(\Theta_0)$. Ao reescrever a equação (2.6), tem-se que a estatística sob as condições de regularidade é dada por

$$\Lambda = -2 \left[\log(L(\hat{\delta}_0, \mathbf{x})) - \sum_{t=1}^T \log(L_t(\hat{\delta}_t, \mathbf{x})) \right],$$

desde que a contribuição da estatística seja priorizada, a partir da soma dos logaritmos das funções de verossimilhanças sobre as T transações, em que se assume uma cadeia *markoviana* de primeira ordem em cada transição (Lara et al., 2018).

A estatística de *Wald* para a t -ésima transição é dada por

$$W_t = [\hat{\delta}_t - \delta_0]^T \hat{\mathcal{F}}_{\hat{\delta}_t} [\hat{\delta}_t - \delta_0], \quad t = 1, 2, \dots, T, \quad (2.7)$$

em que $\hat{\mathcal{F}}_{\hat{\delta}_t}$ é uma matriz avaliada $\hat{\delta}_t$. Matricialmente, a equação (2.7) pode ser reescrita da seguinte forma

$$W = \begin{bmatrix} \hat{\delta}_1 - \delta_0 \\ \hat{\delta}_2 - \delta_0 \\ \vdots \\ \hat{\delta}_T - \delta_0 \end{bmatrix}^T \begin{bmatrix} \hat{\mathcal{F}}_{\hat{\delta}_1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \hat{\mathcal{F}}_{\hat{\delta}_2} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \hat{\mathcal{F}}_{\hat{\delta}_T} \end{bmatrix} \begin{bmatrix} \hat{\delta}_1 - \delta_0 \\ \hat{\delta}_2 - \delta_0 \\ \vdots \\ \hat{\delta}_T - \delta_0 \end{bmatrix} = \sum_{t=1}^T W_t.$$

Assintoticamente, a estatística *Wald* tem distribuição χ_v^2 em que v representa o número de graus de liberdade. A obtenção de v é feita pela diferença entre o número total de parâmetros no modelo envolvendo as T transições e o número de parâmetros do modelo sob estacionariedade. Conforme Lara et al. (2018), um problema em usar essa estatística é que na prática δ_0 é desconhecido, e uma alternativa seria utilizar o valor estimado de $\hat{\delta}_0$ sob a hipótese nula.

Adicionalmente, os autores Lara et al. (2018) desenvolveram estudo por simulação para avaliar a performance dessas estatísticas (razão de verossimilhança e *Wald*) com relação ao clássico (Anderson e Goodman), verificando que, com relação ao erro tipo I, os testes propostos demonstraram um melhor desempenho sobre o teste clássico. Contudo, para grandes amostras, as estatísticas podem ser consideradas assintoticamente equivalentes. Para o erro tipo II, os autores concluíram que com tamanho de amostra 50 o teste clássico foi mais poderoso. Entretanto, com o aumento do tamanho da amostra, os testes também podem ser considerados assintoticamente equivalentes.

Em resumo, os testes propostos por Lara et al. (2018) apresentaram resultados satisfatórios. De forma semelhante, este trabalho tem como proposta avaliar a performance da estatística gradiente com as demais descritas.

2.2.4.1 Estatística gradiente

Equivalente ao trabalho de Lara et al. (2018), o teste para estatística gradiente foi baseado no fato de que os modelos de transição (2.2 e 2.4) são modelos lineares generalizados com estrutura

de dependência no preditor linear. Assim, supondo uma variável resposta do tipo categórica e tempo discreto (pré-definidos), o modelo de transição tem $\mathbf{y}_{it}$ como o vetor das respostas do i -ésimo indivíduo na t -ésima ocasião e $\mathbf{x}_{it}$ vetor das variáveis explicativas. De acordo com as condições de regularidade, o vetor de parâmetros a ser estimado $\boldsymbol{\delta}$ tem distribuição assintoticamente normal, ou seja, $\hat{\boldsymbol{\delta}}_t \sim N(\boldsymbol{\delta}, \mathfrak{S}^{-1})$ e $\mathfrak{S}$ é a matriz de informação de *Fisher*. Então, sob H_0 a estatística gradiente para t -ésima transição é dada por

$$G_t = U(\boldsymbol{\delta}_0)' (\hat{\boldsymbol{\delta}}_t - \boldsymbol{\delta}_0) \quad (2.8)$$

em que $U(\boldsymbol{\delta}_0) = \partial \ell(\boldsymbol{\delta}_0) / \partial \boldsymbol{\delta}_0$ é primeira derivada (função score) do modelo sob processo estacionário, $\hat{\boldsymbol{\delta}}_t$ e $\boldsymbol{\delta}_0$ representam as estimativas dos parâmetros sob processo não estacionário e estacionário, respectivamente. Essa estatística é descrita como mais simples, pois sua expressão não necessita de uma matriz inversa, por exemplo. Assintoticamente, G_t tem distribuição χ^2 com ν graus de liberdade (diferença entre os parâmetros do modelo não estacionário e estacionário), similar às estatísticas clássicas, razão de verossimilhança e *Wald*. Ademais, é possível encontrar no Apêndice (Capítulo 5) as derivadas das duas classes dos modelos de transição (nominal e ordinal).

2.3 Estudo de simulação

Nessa seção, apresenta-se um estudo de simulação seguindo todos os procedimentos implementados por Lara et al. (2018). Os autores utilizaram dados reais (condição de pacientes com problemas respiratórios) de ensaios que, originalmente, contêm cinco ocasiões no tempo, correspondendo a $T = 4$ transições de primeira ordem, a fim de simular dados categóricos estacionários (simulados de uma matriz de transição exclusiva) e não estacionários (de T matrizes de transições distintas), para testar o erro tipo I e o poder do teste das estatísticas Anderson e Goodman, razão de verossimilhança e *wald*, em ambos os processos. Incorporou-se, nesse procedimento, a estatística gradiente com o intuito de confrontá-las.

Para a simulação no caso nominal, utilizaram-se dados sobre o comportamento de forrageio de fêmeas parasitoide *Telenomus podisi* com cinco ocasiões no tempo, a fim de gerar dados categóricos nominais estacionários e não estacionários, para testar o erro tipo I e o poder dos quatro testes descritos. Dez mil amostras foram simuladas com três tamanhos diferentes ($N = 100$, $N = 130$ e $N = 500$) e com três ($T = 2$), quatro ($T = 3$) e cinco ($T = 4$) ocasiões no tempo para cenário I (processo estacionário) e cenário II (processo não estacionário). Definiu-se uma cadeia *markoviana* de primeira ordem nos dois cenários e um efeito de tratamento, em que o preditor linear do modelo é descrito por

$$\eta = \log \left(\frac{\pi_{ab}(t)(\mathbf{x})}{\pi_{ak}(t)(\mathbf{x})} \right) = \lambda_{b(t)} + \beta_{bt}(\text{tratamento}) + \alpha_{bt}(\text{resposta anterior}). \quad (2.9)$$

Ao ajustar o modelo aos dados (2.9) supondo os dois cenários (processo estacionário e não estacionário), construíram-se as matrizes de probabilidades de transição. Consequentemente, o cálculo das estatísticas de Anderson e Goodman (1957), Razão de verossimilhança, *wald* e gradiente (interesse de trabalho) foram obtidas. Nos dois cenários, as taxas de rejeição para cada estatística foram obtidas para os níveis de significância 1%, 5% e 10%.

Para o caso ordinal, foram simuladas 10 mil amostras para os tamanhos ($N = 70$, $N = 100$ e $N = 500$) e três, quatro e cinco ocasiões no tempo para o cenário I (processo estacionário) e cenário II (processo não estacionário). O preditor linear do modelo para esse caso é descrito por

$$\eta = \log \left(\frac{\gamma_{ab}(t)}{1 - \gamma_{ab}(t)} \right) = \lambda_{b(t)} + \beta_t(\text{tratamento}) + \alpha_t(\text{resposta anterior}). \quad (2.10)$$

De acordo com Agresti (2002), variáveis nominais podem ser estudadas como ordinais sem considerar a ordem. Dessa forma, a implementação computacional foi feita no *software* R (R Core Team, 2020), usando o pacote ordinal (Christensen, 2018), para ajustar os dois modelos. Utilizou-se desse pacote nos modelos nominais pelo fato da função *clm* ter a função gradiente implementada e por retornar probabilidades estimadas similares da função multinom (não tem a função gradiente implementada) do pacote *nnet*. Além do mais, o pacote *markovchain* (Spedicato et al., 2016) foi utilizado para auxiliar no processo de simulação dos dados.

2.3.1 Dados categóricos nominais

Com relação aos resultados da simulação para o erro tipo I (Tabela 2.1), verificou-se que as taxas de rejeições das estatísticas estavam em concordância com os níveis de significância (10%, 5% e 1%). Contudo, o que se mostrou ser mais conservador foi o teste clássico (Anderson e Goodman). Com o aumento do tamanho de amostra, as taxas de rejeição diminuíram e aproximavam-se da significância desejada. Além disso, o número de ocasiões do tempo (T) não mostrou influenciar nos resultados das estatísticas. Em particular, a estatística gradiente mostrou bom desempenho principalmente com o aumento da amostra, em que todas as estatísticas mostraram ser assintoticamente equivalentes.

Tabela 2.1: Taxas de rejeição (erro tipo I) para os testes de Anderson e Goodman (AG), Wald (W), razão de verossimilhança (LR) e Gradiente (G), decorrente de 10.000 simulações assumindo um modelo estacionário nominal de primeira ordem.

Tempo Significância		$T = 3$			$T = 4$			$T = 5$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
$N = 100$	Teste									
	AG	0,1000	0,0494	0,0080	0,0909	0,0438	0,0081	0,0845	0,0392	0,0063
	W	0,1079	0,0559	0,0126	0,1141	0,0622	0,0181	0,1069	0,0573	0,0158
	LR	0,1028	0,0529	0,0096	0,0956	0,0462	0,0081	0,0893	0,0408	0,0053
	G	0,1147	0,0618	0,0141	0,1181	0,0622	0,0151	0,1170	0,0615	0,0122
$N = 130$	Teste									
	AG	0,1032	0,0540	0,0100	0,0992	0,0472	0,0096	0,1008	0,0519	0,0099
	W	0,1125	0,0609	0,0151	0,1219	0,066	0,0205	0,1176	0,0685	0,0224
	LR	0,1096	0,0563	0,0115	0,1033	0,0502	0,0109	0,1048	0,0555	0,0109
	G	0,1172	0,0640	0,0161	0,123	0,0663	0,0151	0,1276	0,0709	0,0185
$N = 500$	Teste									
	AG	0,1045	0,0524	0,0126	0,1047	0,0555	0,0126	0,1014	0,0519	0,0104
	W	0,1057	0,0544	0,0133	0,1098	0,0580	0,0147	0,1045	0,0544	0,0112
	LR	0,1062	0,0533	0,0128	0,1072	0,0571	0,0134	0,1032	0,0534	0,0104
	G	0,1085	0,0549	0,0133	0,1114	0,0593	0,0139	0,1073	0,0586	0,0121

Analisando o poder do teste (Tabela 2.2), observou-se que o poder aumenta com o tamanho da amostra e com o número de ocasiões do tempo (T). Verifica-se que o teste proposto gradiente retornou um poder acima dos demais, para $N = 100$ e $N = 130$. Com um tamanho $N = 500$, todas as estatísticas são consideradas assintoticamente equivalentes.

Tabela 2.2: Taxas de rejeição (poder do teste) para o teste de Anderson e Goodman (AG), Wald (W), razão de verossimilhança (LR) e Gradiente (G), decorrente de 10.000 simulações assumindo um modelo não estacionário nominal de primeira ordem.

Tempo	Significância	$T = 3$			$T = 4$			$T = 5$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
$N = 100$	Teste									
	AG	0,6341	0,5026	0,2523	0,6814	0,5532	0,3101	0,8030	0,6840	0,4270
	W	0,6118	0,4694	0,2140	0,6653	0,5349	0,2881	0,7650	0,6610	0,4020
	LR	0,6530	0,5272	0,2809	0,6866	0,5644	0,3242	0,8160	0,7000	0,4540
	G	0,6769	0,5596	0,3225	0,7125	0,5978	0,3668	0,8510	0,7550	0,5290
$N = 130$	Teste									
	AG	0,7703	0,6614	0,4130	0,8362	0,7427	0,5161	0,9222	0,8595	0,6747
	W	0,7546	0,6356	0,3772	0,8323	0,7313	0,4934	0,9129	0,8458	0,6455
	LR	0,7846	0,6812	0,4408	0,8416	0,7503	0,5250	0,9301	0,8717	0,6990
	G	0,7988	0,7046	0,4804	0,8550	0,7685	0,5583	0,9402	0,8938	0,7422
$N = 500$	Teste									
	AG	1,000	0,9998	0,9984	1,000	1,000	0,9999	1,000	1,000	1,000
	W	1,000	0,9997	0,9982	1,000	1,000	0,9999	1,000	1,000	1,000
	LR	1,000	0,9998	0,9985	1,000	1,000	0,9999	1,000	1,000	1,000
	G	1,000	0,9998	0,9986	1,000	1,000	0,9999	1,000	1,000	1,000

2.3.2 Dados categóricos ordinais

Com relação aos resultados do erro tipo I (Tabela 2.3), identificou-se que as taxas de rejeições das estatísticas de Anderson e Goodman, razão de verossimilhança e gradiente retornaram valores maiores que 10%, 5% e 1% para $N = 70$ e $N = 100$. Já o teste de *Wald* mostrou-se ser mais conservador, pois suas taxas estavam próximas da desejável. Para $N = 500$, todas as estatísticas mostraram ser assintoticamente equivalentes, pois as taxas de rejeição de todas as estatísticas aproximaram-se da significância pretendida. Além disso, o número de ocasiões do tempo (T) não mostrou influenciar nos resultados. Particularmente, a estatística gradiente não apresentou bom resultado ao obter as taxas de rejeição do erro tipo I para o modelo com dados ordinais. Uma vez que, essas taxas foram maiores que a significância desejada.

Tabela 2.3: Taxas de rejeição (erro tipo I) para os testes de Anderson e Goodman (AG), Wald (W), razão de verossimilhança (LR) e Gradiente (G), decorrente de 10.000 simulações assumindo um modelo estacionário ordinal de primeira ordem.

Tempo	Significância	$T = 3$			$T = 4$			$T = 5$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
$N = 70$	Teste									
	AG	0,1348	0,0732	0,0166	0,1380	0,0768	0,0191	0,1357	0,0716	0,0188
	W	0,0959	0,0457	0,0081	0,0849	0,0398	0,0080	0,0663	0,0321	0,0060
	LR	0,1251	0,0668	0,0138	0,1176	0,0636	0,0130	0,1052	0,0527	0,0100
	G	0,1478	0,0848	0,0218	0,1543	0,0877	0,0238	0,1499	0,0843	0,0228
$N = 100$	Teste									
	AG	0,1286	0,0703	0,0154	0,1345	0,0709	0,0155	0,1346	0,0734	0,0181
	W	0,1036	0,0507	0,0107	0,0933	0,0436	0,0079	0,0841	0,0407	0,0086
	LR	0,1251	0,0679	0,0154	0,1258	0,0621	0,0113	0,1166	0,0603	0,0120
	G	0,1397	0,0791	0,0197	0,1508	0,0839	0,0204	0,1554	0,0865	0,0206
$N = 500$	Teste									
	AG	0,1040	0,0545	0,0116	0,1066	0,0562	0,0130	0,1138	0,0564	0,0122
	W	0,1000	0,0518	0,0102	0,0985	0,0516	0,0115	0,1028	0,0502	0,0089
	LR	0,1038	0,0551	0,0115	0,1048	0,0541	0,0125	0,1104	0,0552	0,0112
	G	0,1060	0,0563	0,0130	0,1104	0,0574	0,0131	0,1181	0,0601	0,0126

Para as taxas do poder do teste (Tabela 2.4), foi observado que aos níveis de significância 10%, 5% e 1%, as taxas aumentam com o tamanho da amostra e com o número de ocasiões do tempo (T). Com relação ao teste proposto (Gradiente) observou-se um poder acima dos demais (AG, RV e *Wald*), embora assintoticamente todos sejam similares. Esses resultados corroboram com a simulação desenvolvida por

Lara et al. (2018), que ainda indicam o teste de *Wald* como o menos poderoso.

Tabela 2.4: Taxas de rejeição (poder do teste) para o teste de Anderson e Goodman (AG), Wald (W), razão de verossimilhança (LR) e Gradiente (G), decorrente de 10.000 simulações assumindo um modelo não estacionário ordinal de primeira ordem.

Tempo Significância		$T = 3$			$T = 4$			$T = 5$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
$N = 70$	Teste									
	AG	0,5156	0,3722	0,1600	0,8319	0,7308	0,4920	0,9353	0,8827	0,7137
	W	0,4051	0,2583	0,0765	0,7269	0,5825	0,2948	0,8530	0,7483	0,4758
	LR	0,5134	0,3695	0,1543	0,8163	0,7083	0,4530	0,9203	0,8566	0,6616
	G	0,5607	0,4238	0,2072	0,8526	0,7614	0,5306	0,9427	0,8971	0,7434
$N = 100$	Teste									
	AG	0,7287	0,5958	0,3263	0,9626	0,9275	0,7918	0,9937	0,9870	0,9491
	W	0,6457	0,4817	0,2047	0,9344	0,8737	0,6545	0,9859	0,9677	0,8756
	LR	0,7369	0,6014	0,3317	0,9615	0,9232	0,7819	0,9931	0,9849	0,9391
	G	0,7685	0,6474	0,3927	0,9681	0,9405	0,8273	0,9952	0,9891	0,9588
$N = 500$	Teste									
	AG	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	W	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	LR	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	G	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

A partir das simulação para dados categóricos nominais e ordinais, salienta-se que estudos com dados longitudinais e tamanhos amostrais pequenos podem levar a problemas de convergência ao ajustar o modelo.

2.4 Estudo na Entomologia

Como motivação, apresenta-se um experimento desenvolvido e adaptado especificamente para este trabalho. O estudo foi realizado no Laboratório de biologia de insetos da Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo. O objetivo foi estudar o comportamento de forrageio de fêmeas parasitoides *Telenomus podisi*, um parasitoide potencial na área agrícola, quando exposta a ovos de *Euchistus heros* (sadio ou previamente inoculados pela própria espécie). Deseja-se estudar o potencial da fêmea *Telenomus podisi* no reconhecimento de ovos previamente parasitados, isto é, se a mesma evita competição intraespécie.

Nesse experimento, as fêmeas foram deixadas por 24 horas acasalando com machos da espécie e, em seguida, transferidas para placas de petri que foram aleatorizados para conter quatro ovos não parasitados (sp0) ou previamente parasitados (sp1). Foram preparados 16 placas de petri para cada tratamento (sp0 e sp1), controlando as condições de luz e clima da sala. Após serem introduzidas nas placas, observaram-se as microvespas qualificando-as nas categorias: 1 - voar ou caminhar, 2 - tamborilar e 3 - ovipositar ou marcar. Os comportamentos foram analisados em 5 ocasiões de tempo (2, 4, 6, 8 e 10 minutos), caracterizando um estudo longitudinal de tempo discreto com variável resposta qualitativa nominal.

No total foram observados 32 insetos, sendo 16 insetos de cada tratamento. Contudo, por se tratar de uma amostra pequena para estudos longitudinais, foi preciso simular mais observações com o objetivo de não ter problemas de convergência ao ajustar o modelo aos dados. De acordo com os tamanhos amostrais utilizados na seção 2.3.1, observou-se que um valor próximo de 100 seria o ideal. Assim, foram simulados 64 insetos, 32 para cada tratamento. Totalizando um cenário experimental com 96 observações, 48 em cada tratamento.

2.4.1 Resultados

A Figura 2.1 descreve a frequência total observada do comportamento da parasitoide nas categorias 1 - voar ou caminhar, 2 - tamborilar e 3 - ovipositar ou marcar, para cada tempo observado nos grupos com ovos não parasitados (sp0) e previamente parasitados por *Telenomus podisi* (sp1). Observou-se um total de 48 parasitoides, em cada grupo e, ao analisar o comportamento das categorias, há indícios de efeito de tratamento.

Figura 2.1: Frequências de comportamentos (1, 2 ou 3) da microvespa fêmea *Telenomus podisi*, em ovos não parasitados e previamente parasitados, para cada tempo observado.

Ao ajustar o modelo de transição de ordem um sob estacionariedade, verificou-se que o efeito de tratamento foi significativo (valor-p < 0,0001).

Quanto à verificação de estacionariedade do processo, ao longo do tempo, modelos compostos por tratamento (ovos *Euchistus heros*) e resposta anterior (dependência de primeira ordem de *markov*) foram ajustados considerando processo estacionário e não estacionário. Os resultados das estatísticas e valor-p em parêntese de Anderson e Goodman, *Wald*, razão de verossimilhança e gradiente foram, respectivamente, $\xi = 13,6736$ (0,9536), $W = 11,6100$ (0,9840), $\Lambda = 11,6301$ (0,9837) e $G = 11,8852$ (0,9811) com 24 graus de liberdade. Todas essas estatísticas não foram significativas e indicam aplicação de um modelo estacionário. Analisando a Tabela 2.5, tem-se que todas as covariáveis foram significativas para o modelo e retornou erros padrões baixos. A fim de obter as probabilidades de transição (Figura 2.2), utilizaram-se as estimativas dos parâmetros.

Tabela 2.5: Valores das estimativas dos parâmetros, erro padrão e valor- p do modelo de transição assumindo estacionariedade para os dados de comportamento parasitoide.

Parâmetros	Estimativas	Erro padrão	Valor- p
Interceptos			
λ_{a2}	-0,3050	0,2176	
λ_{a3}	0,4979	0,2278	
Covariáveis			
$\beta_2(sp1)$	1,4199	0,2438	<0,0001
$\beta_3(sp1)$	1,9962	0,2920	<0,0001
$\alpha_2(resposta\ anterior\ 2)$	-0,9525	0,3134	0,0024
$\alpha_3(resposta\ anterior\ 2)$	-0,9612	0,3429	0,0051
$\alpha_2(resposta\ anterior\ 3)$	-1,1155	0,2836	0,0001
$\alpha_3(resposta\ anterior\ 3)$	-1,1628	0,2849	<0,0001

Com base nos resultados da Figura 2.2, referentes às estimativas das probabilidades de mudanças dos estados nos dois tratamentos, pode-se afirmar que os parasitoides liberados na placa com ovos não parasitados têm maior probabilidades de oviposição ou marcação do que os liberados em ovos previamente parasitados. A probabilidade de um parasitoide sair do estado voar/caminhar ou tamborilar para ovipositar/marcar é 0,3780 e 0,6138 em ovos não parasitados, enquanto, em ovos previamente parasitados essas probabilidades, são respectivamente, 0,0763 e 0,1776. Assim, tem-se que a chance das fêmeas parasitoides *Telenomus podisi* ovipositar/marcar saindo dos estados voar/caminhar ou tamborilar é 7,35 vezes maior no tratamento sp0.

Figura 2.2: Estimativas das probabilidades assumindo estacionariedade para a primeira transição dos dados de comportamento parasitoide.

2.5 Conclusão

Modelos de transição são essenciais para descrever as probabilidades das mudanças de uma variável resposta categórica, assim como verificar a existência de efeitos de covariáveis. A condição de estacionariedade é desejável nesses modelos, mas nem sempre é satisfeita. Ao propor o teste gradiente como alternativa aos de Anderson e Goodman, Wald e razão de verossimilhança, a mesma obteve resultados interessantes. Para dados nominais, o erro tipo I e poder do teste retornou bom desempenho. Diferente do caso ordinal, em que os resultados não foram bons para o erro tipo I.

A utilização dessa metodologia possibilitou ainda a análise de uma situação típica em estudos entomológicos, cujo interesse era verificar os diferentes comportamentos durante o forrageio por ovos de fêmeas parasitoides *Telenomus podisi* com efeito de ovos *Euchistus heros*. Os resultados mostraram que

há diferença entre o número de eventos de oviposição ou marcação em ovos não parasitados e previamente parasitados. De acordo com as estatísticas descritas neste trabalho, o processo é estacionário. Ajustando um modelo de transição com essas características e estimando as probabilidades das mudanças, tem-se que microvespa fêmea *Telenomus podisi* é mais propícia à ovipositar ou marcar em ovos não parasitados do que em ovos previamente parasitados, evitando a competição intraespécie em *Telenomus podisi*.

Referências

- Agresti, A. (2002). Categorical data analysis 2nd edition. *Edited by: Wiley JSI. Hoboken, NJ*, pages 228–230.
- Albert, P. S. (2000). A transitional model for longitudinal binary data subject to nonignorable missing data. *Biometrics*, 56(2):602–608.
- Anderson, T. W. e Goodman, L. A. (1957). Statistical inference about markov chains. *The Annals of Mathematical Statistics*, pages 89–110.
- Christensen, R. H. B. (2018). Cumulative link models for ordinal regression with the r package ordinal. *Accessed in 2020*.
- Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Heagerty, P. J., Zeger, S., et al. (2002). *Analysis of longitudinal data*. Oxford University Press.
- Ferrari, S. e Pinheiro, E. C. (2014). Small-sample likelihood inference in extreme-value regression models. *Journal of Statistical Computation and Simulation*, 84(3):582–595.
- Ganjali, M. e Rezaee, Z. (2007). A transition model for analysis of repeated measure ordinal response data to identify the effects of different treatments. *Drug information journal*, 41(4):527–534.
- Heagerty, P. J. (2002). Marginalized transition models and likelihood inference for longitudinal categorical data. *Biometrics*, 58(2):342–351.
- Hoel, P. G., Port, S. C., e Stone, C. J. (1986). *Introduction to stochastic processes*. Waveland Press.
- Hosmer, D. W., Lemeshow, S., e Sturdivant, R. X. (2013). *Applied logistic regression*, volume 398. John Wiley & Sons.
- Karlin, S. e Taylor, H. (1975). *A first course in stochastic processes*. Academic Press, New York.
- Lai, C. K. (1967). *Markov chains with stationary transition probabilities*. Springer-Verlag.
- Lara, I., Hinde, J., de Castro, A., e da Silva, I. (2017). A proportional odds transition model for ordinal responses with an application to pig behaviour. *Journal of Applied Statistics*, 44(6):1031–1046.
- Lara, I. A. R., Hinde, J., e Taconeli, C. A. (2018). An alternative method for evaluating stationarity in transition models. *Journal of Statistical Computation and Simulation*, 87(15):2962–2980.
- Lee, K. e Daniels, M. J. (2007). A class of markov models for longitudinal ordinal data. *Biometrics*, 63(4):1060–1067.
- Lee, T., Judge, G., e Takayama, T. (1965). On estimating the transition probabilities of a markov process. *Journal of Farm Economics*, 47(3):742–762.
- Lemonte, A. (2016). *The gradient test: another likelihood-based test*. Academic Press.

- Lemonte, A. J. e Ferrari, S. L. (2012). The local power of the gradient test. *Annals of the Institute of Statistical Mathematics*, 64(2):373–381.
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127.
- Molenberghs, G. e Verbeke, G. (2006). *Models for discrete longitudinal data*. Springer Science & Business Media.
- Peterson, B. e Harrell Jr, F. E. (1990). Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 39(2):205–217.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ross, S. M., Kelly, J. J., Sullivan, R. J., Perry, W. J., Mercer, D., Davis, R. M., Washburn, T. D., Sager, E. V., Boyce, J. B., e Bristow, V. L. (1996). *Stochastic processes*, volume 2. Wiley New York.
- Spedicato, G. A., Kang, T. S., Yalamanchi, S. B., Yadav, D., e Cordon, I. (2016). The markovchain package: a package for easily handling discrete markov chains in r. *Accessed in 2020*.
- Stram, D. O., Wei, L. J., e Ware, J. H. (1988). Analysis of repeated ordered categorical outcomes with possibly missing observations and time-dependent covariates. *Journal of the American Statistical Association*, 83(403):631–637.
- Terrell, G. R. (2002). The gradient statistic. *Computing Science and Statistics*, 34(34):206–215.
- Vargas, T. M., Ferrari, S. L., e Lemonte, A. J. (2014). Improved likelihood inference in generalized linear models. *Computational Statistics & Data Analysis*, 74:110–124.
- Ware, J. H., Lipsitz, S., e Speizer, F. E. (1988). Issues in the analysis of repeated categorical outcomes. *Statistics in Medicine*, 7(1-2):95–107.
- Williams, O. D. e Grizzle, J. E. (1972). Analysis of contingency tables having ordered response categories. *Journal of the American Statistical Association*, 67(337):55–63.
- Yang, X., Shoptaw, S., Nie, K., Liu, J., e Belin, T. R. (2007). Markov transition models for binary repeated measures with ignorable and nonignorable missing values. *Statistical Methods in Medical Research*, 16(4):347–364.
- Yu, F., Morgenstern, H., Hurwitz, E., e Berlin, T. R. (2003). Use of a markov transition model to analyse longitudinal low-back pain data. *Statistical Methods in Medical Research*, 12(4):321–331.
- Zeng, L. e Cook, R. J. (2007). Transition models for multivariate longitudinal binary data. *Journal of the American Statistical Association*, 102(477):211–223.

3 MODELOS DE MULTIESTADOS NO ESTUDO COMPORTAMENTAL DO ÁCARO PREDADOR *PHYTOSEIULUS PERSIMILIS* E DA MICROVESPA *TELENOMUS PODISI* PARA CONTROLE BIOLÓGICO DE PRAGAS

Resumo

O controle biológico de pragas por meio de inimigos naturais (como por exemplo fungos patogênicos e parasitoides) são formas sustentáveis para a proteção de plantas e, consequentemente da produção agrícola. Nesse contexto, tem-se como motivação dois experimentos entomológicos, em que os comportamentos de fêmeas do ácaro predador *Phytoseiulus persimilis* e de fêmeas do parasitoide *Telenomus podisi* foram qualificados no tempo, considerando efeitos potenciais de tratamentos para cada ensaio. Os experimentos caracterizaram estudos longitudinais de tempo contínuo e com variável resposta qualitativa nominal, agrupadas em quatro e cinco categorias mutuamente exclusivas, respectivamente, para as microvespas *Telenomus podisi* e ácaro predador *Phytoseiulus persimilis*. Os modelos de transição de tempo contínuo, também conhecido como modelos multiestados, são uma alternativa viável para analisar dados com essas características. Sendo assim, considerou-se nesse trabalho, essa classe de modelos para estudar os comportamentos da microvespa (1 - voar ou caminhar, 2 - tamborilar, 3 - ovipositar e 4 - marcar) e do ácaro predador (1 - procurando presa, 2 - encontrou presa, 3 - alimentando-se, 4 - andando fora da folha e 5 - andando na ponte de *parafilm*), por seus tempos médios de permanência, em cada uma das categorias de resposta, bem como suas probabilidades de preferências. Como resultado, verificou-se efeito significativo de tratamento em ambos os experimentos, ovos não parasitados e previamente parasitados (para microvespa) e *Metarhizium robertsii*, *Beauveria bassiana* e *Beauveria bassiana* x *Metarhizium robertsii* (para ácaro predador). No estudo do ácaro predador, a fêmea *Phytoseiulus persimilis* tem preferência por alimentar-se quando submetido ao composto *Beauveria bassiana* x *Metarhizium robertsii*. Já no estudo que tinha por objetivo estudar o comportamento de forrageio de fêmeas parasitoides *Telenomus podisi*, ao ter contato com o tratamento de ovos previamente parasitados, a microvespa tem preferência pelas categorias de movimento que são voo caminhando e tamborilando. Em contrapartida, quando em contato com o tratamento ovos não parasitados, a microvespa fêmea *Telenomus podisi* tem preferência pela oviposição e marcação. A metodologia aplicada mostrou-se adequada para as duas situações experimentais típicas em estudos entomológicos associados à descrição do comportamento de um ácaro predador ou parasitoide.

Palavras-chave: Entomologia. Dados categorizados longitudinais. Máxima verossimilhança. Intensidades de transição.

3.1 Introdução

Na área agronômica, é crescente o número de estudos que envolvem medidas preventivas para a proteção das culturas agrícolas. Cuidados com a quantidade de luz solar, água, solo, ventos e controle de pragas são de suma importância para o bem-estar das plantas. Em particular, o controle biológico é uma abordagem segura e sustentável que combate pragas utilizando inimigos naturais e de custo benefício favorável. Conforme Gámez-Virués et al. (2009), existe um aumento mundial significativo em defesa do controle biológico por meio do uso de parasitoides como uma estratégia eficaz, uma vez que elimina a procriação de pragas que atacam as plantas sem causar danos ao meio ambiente. Ademais, Behie et al. (2015) descrevem que os fungos (entomopatogênicos ou endofíticos) caracterizam papéis importantes na ecologia vegetal, evolução da planta e interações do solo (rizosfera), podendo causar aumento no crescimento das plantas e protegê-las contra as pragas.

Nesse contexto, experimentos entomológicos são delineados para estudar e identificar quais

inimigos naturais têm melhor desempenho na proteção de uma determinada espécie de planta. Em algumas situações, as vespas parasitoides foram usadas no controle biológico de pragas, incluindo a diversidade das espécies e diferentes culturas de pragas (Wang et al., 2019). Em pequenas fazendas ou jardins, o uso de vespas sociais indicou ser um agente competente ao controle de pragas (Prezoto et al., 2019). Com relação aos fungos (entomopatógenos ou endófitos), Jaber e Enkerli (2016) obtiveram resultados significativos, quanto ao crescimento e colonização de favas, utilizando *Beauveria bassiana* e *Metarhizium brunneum*. Em Canassa et al. (2019), os fungos *Metarhizium robertsii* e *Beauveria bassiana* mostraram bons resultados com relação à capacidade endofítica e crescimento em plantas de morango.

Em grande parte dos casos, os estudos qualificam o comportamento de um determinado indivíduo (inseto), em um período de tempo e categorias (nominais ou ordinais) mutuamente exclusivas, tecnicamente chamadas de estados dos indivíduos. A análise estatística desses dados requer métodos que levem em conta o processo tempo-estado. Entre os possíveis modelos, destaca-se o modelo de transição, que é uma técnica capaz de analisar dados categorizados no tempo. Ware et al. (1988) classificam esse modelo como condicional por levar em conta as respostas passadas (história do indivíduo) e um conjunto de covariáveis. Molenberghs e Verbeke (2006) definem essa classe como uma extensão dos modelos lineares generalizados, sendo úteis para descrever as mudanças entre as categorias dos indivíduos.

Conforme Lara et al. (2018), o modelo de transição é baseado em processos estocásticos, podendo ser de tempo discreto ou contínuo. Quando as observações dos indivíduos são coletadas, em tempos pré-definidos, define-se como sendo discreto. Um exemplo é o estudo da gravidade de lesões em suínos, quando submetidos a duas condições de criação: com e sem enriquecimento ambiental (Lara et al., 2017). Já, quando os valores são coletados em um intervalo de tempo (não pré-definido), tem-se o processo de tempo contínuo. Oliveira et al. (2013, 2018) e Meira-Machado et al. (2009), descrevem esses modelos para analisar a condição física dos pacientes sobre uma determinada doença.

O modelo de transição de tempo contínuo também é conhecido como modelo multiestados. Em adição ao modelo de tempo discreto, nesta classe, podem-se estimar as intensidades de transição, que são parâmetros associados às probabilidades de transições, fundamentais para descrever simultaneamente as mudanças de comportamento dos indivíduos, assim como a “velocidade” com que isso ocorre no tempo (Kalbfleisch e Lawless, 1985). De acordo com Gimenez et al. (2003), o modelo multiestados é uma metodologia poderosa, quando se tem interesse em questões biológicas relacionadas à dispersão e/ou variedade individual em populações silvestres, como bem ilustrado em Conn e Cooch (2009), por meio de um estudo de conjuntividade em pássaros *Haemorrhous mexicanus*. Ao capturar e recapturar os pássaros, observam-se os estados dessa doença e suas taxas de mudanças no tempo.

Nessa classe de modelos, também é de interesse o tempo até a ocorrência do evento, sendo que pode ser observado ou censurado. A censura é descrita no momento em que o indivíduo tem seu tempo até a ocorrência interrompido (Klein e Moeschberger, 2006). A existência dessa observação censurada ocorre por diversas causas, podendo estar relacionada ou não com o evento de interesse. Os tempos censurados devem ser usados na análise estatística, pois, mesmo sendo incompletos, fornecem informação sobre o tempo de falha dos indivíduos e a não utilização levará a resultados imprecisos. Além disso, os modelos de multiestados podem ser considerados como uma generalização dos modelos de sobrevivência, visto que Meira-Machado et al. (2009) descrevem o modelo de mortalidade como o mais simples dos multiestados para análise de sobrevivência.

Em face da importância do controle biológico de pragas e da aplicabilidade do modelo de transição, o presente trabalho tem como objetivo estender essa metodologia à entomologia, tendo como motivação dois experimentos desenvolvidos na Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo. O primeiro experimento tem como motivação verificar fêmeas do ácaro predador *Phytoseiulus persimilis* em plantas de feijão inoculadas com os tratamentos fúngicos (*Metarhizium robertsii*, *Beauveria bassiana* e *Beauveria bassiana* x *Metarhizium robertsii*). O segundo tem como motivação

verificar a habilidade de fêmeas do parasitoide *Telenomus podisi* em reconhecer ovos do percevejo *Euschistus heros* (não parasitados ou previamente parasitados).

Esse capítulo está estruturado da seguinte forma: na seção 3.2 faz-se uma breve revisão de processo estocástico de tempo contínuo e do modelo multi-estados. Na seção 3.3, apresentam-se os estudos de motivação e os métodos a serem empregados na análise dos dados. Na seção 3.4, apresentam-se os resultados das análises dos experimentos e a discussão dos mesmos. Finalmente, na seção 3.5 tem-se as considerações finais desse capítulo.

3.2 Processos estocásticos e modelos multiestados

Um processo estocástico é descrito como um fenômeno aleatório que varia, ao longo do tempo e tem características imprevisíveis, já que pode produzir resultados diferentes para cada vez que o processo é executado. Segundo Karlin e Taylor (1975), um processo estocástico envolve relações entre variáveis aleatórias e é uma sequência de eventos regidos por leis probabilísticas podendo ser de tempo discreto ou contínuo. Nesse contexto, considere indivíduos movendo-se livremente por k categorias no tempo. Um processo estocástico pode ser denotado pela variável aleatória Y_t , assumindo valores no conjunto $S = \{1, 2, \dots, k\}$, chamado de espaço de estados do processo. O parâmetro t , por sua vez, assume valores no conjunto τ , que se referem ao tempo do processo: se enumerável define o processo de tempo discreto, caso contrário, tem-se o processo estocástico de tempo contínuo. Contudo, na maioria das aplicações da entomologia, o processo estocástico associado ao experimento é de tempo contínuo e, por isso, é objeto de estudo deste trabalho.

3.2.1 Cadeias de Markov de tempo contínuo

Seja $Y_t \in S$ um processo estocástico de tempo contínuo, $\tau = \{0 < l < t\}$, as probabilidades de transição do estado a para o estado b são definidas por

$$\pi_{ab}(l, t) = P(Y_t = b | Y_l = a), \quad \forall a, b \in S.$$

Matricialmente, tem-se:

$$\mathbf{P}(l, t) = \begin{pmatrix} \pi_{11}(l, t) & \pi_{12}(l, t) & \dots & \pi_{1k}(l, t) \\ \pi_{21}(l, t) & \pi_{22}(l, t) & \dots & \pi_{2k}(l, t) \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{k1}(l, t) & \pi_{k2}(l, t) & \dots & \pi_{kk}(l, t) \end{pmatrix}.$$

De acordo com Lara et al. (2017), em um processo *markoviano*, as probabilidades podem ser classificadas como homogêneas, ao longo do tempo, o que define um processo estacionário. Para o caso contínuo, tem-se:

$$\pi_{ab}(t) = P(Y_{t+l} = b | Y_l = a),$$

ou ainda, em termos matriciais $\mathbf{P}(t) = \mathbf{P}(l, l + t)$. A matriz de transição $\mathbf{P}(t)$ tem elementos que satisfazem, $\pi_{ab}(t) \geq 0$, $a, b = 1, 2, \dots, k$ e $\sum_{b=1}^k \pi_{ab}(t) = 1$. Sob essa condição, Karlin e Taylor

(1975) mostram que, se todos os estados $y \in S$ são não absorventes (dado a ocorrência de um estado, a probabilidade dele ocorrer novamente é diferente de zero), então deve existir uma função de distribuição tal que:

$$\frac{1 - F_y(t + l)}{1 - F_y(l)} = 1 - F_y(t). \quad (3.1)$$

sendo uma distribuição exponencial, com parâmetro q_y , e satisfaz plenamente a equação (3.1), ou seja:

$$f_y(t) = q_y \exp(-q_y t) \quad t \geq 0,$$

em que $\mu = \frac{1}{q_y}$ e $\sigma^2 = \frac{1}{q_y^2}$ correspondem, respectivamente, à média e à variância da distribuição. Caso y seja absorvente, tem-se $q_y = 0$. Assim as probabilidades de transição estacionárias podem ser descritas como função da distribuição exponencial e, além disso, assume-se que são contínuas e diferenciáveis no tempo (Hoel et al., 1972). Tem-se

$$\left[\frac{\partial \pi_{ab}(t)}{\partial t} \right]_{t=0} = q_{ab}(t),$$

que são as intensidades de transição do processo *markoviano*. Demonstra-se ainda que:

$$q_{ab}(t) = \lim_{\Delta_t \rightarrow 0} \frac{\pi_{ab}(t, t + \Delta_t) - \pi_{ab}(t)}{\Delta_t}.$$

As intensidades podem ser interpretadas como taxas de transição do processo no tempo, e satisfazem, $q_{aa} \leq 0$ com $a = 1, 2, \dots, k$, $q_{ab} \geq 0$ para $a \neq b$ e $\sum_{b=1}^k q_{ab}(t) = 0$. Matricialmente, tem-se:

$$\mathbf{Q}(t) = \begin{pmatrix} q_{11}(t) & q_{12}(t) & \dots & q_{1k}(t) \\ q_{21}(t) & q_{22}(t) & \dots & q_{2k}(t) \\ \vdots & \vdots & \ddots & \vdots \\ q_{k1}(t) & q_{k2}(t) & \dots & q_{kk}(t) \end{pmatrix}.$$

em que, para um processo estacionário, $\mathbf{Q}(t) = \mathbf{Q}(l, l + t)$. De acordo com Singer e Spilerman (1976), as intensidades estão relacionadas às probabilidades de transição e ao tempo, da seguinte forma: $\pi_{ab} = -\frac{q_{ab}}{q_{aa}}$ é a probabilidade de uma observação que está no estado a se mover para o estado b , dado a ocorrência de uma transição; $-\frac{1}{q_{aa}}$ é o tempo esperado de uma observação que está no estado a permanecer nesse mesmo estado.

3.2.2 Modelo multiestado

O modelo multiestado é uma generalização do processo estocástico de tempo contínuo descrito na seção 3.2.1 para se incorporar covariáveis na análise. Isso é possível, por meio de modelos de regressão,

para os quais se supõe uma função de distribuição de probabilidade para o tempo, sendo a distribuição exponencial a mais usual pela relação de identidade (3.1). Porém seu uso não é restrito, e as distribuições *Weibull* e log-normal são outras opções. Considere uma amostra aleatória de indivíduos com $i = 1, 2, \dots, n$ com variável resposta inerente a processo estocástico $Y_t \in S = \{1, 2, \dots, k\}, t \geq 0$. Considere também $\pi_{ab}(t)$ a probabilidade de transição da categoria a para a b e $q_{ab}(t)$ a intensidade com que essa mudança acontece. De acordo com Wreede et al. (2010), o modelo de Cox multiestado é o modelo de regressão mais conhecido e utilizado em estudos de dados de sobrevivência, descrito da seguinte forma

$$q_a(t_a, \mathbf{x}_a) = q_a^{(0)}(t_a) \exp\{\boldsymbol{\beta}_a^\top \mathbf{x}_a\} \quad (3.2)$$

em que $a \in \{1, \dots, k\}$ corresponde ao estado das transições, $q_a^{(0)}(t_a)$ o risco ou taxa inicial para a transição a , $\boldsymbol{\beta}_a$ é o vetor de coeficientes de regressão de cada transição a , $\mathbf{x}_a$ o vetor de covariáveis, podendo ser específico para cada transição ou comum a todas as transições. A contribuição individual para o modelo (3.2) é dada por

$$q_{ai}(t_{ai}, \mathbf{x}_{ai}) = q_{ai}^{(0)}(t_{ai}) \exp\{\boldsymbol{\beta}_a^\top \mathbf{x}_{ai}(t_{ai})\}.$$

Putter (2011) e Wreede et al. (2011) descrevem que $\mathbf{x}_{ai}(t_{ai})$ permite que cada transição tenha seu próprio conjunto de covariáveis ou que cada covariável, em cada transição tenha diferentes coeficientes. Ademais Rotolo (2013) descreve que a proporcionalidade das taxas das diferentes transições pode ser especificada pelo uso da taxa inicial e pela adição de uma variável dummy em $\mathbf{x}_{ai}(t_{ai})$. Consequentemente, a função de verossimilhança do modelo (3.2) é dada por

$$L(\boldsymbol{\beta}, \mathbf{q}^{(0)}(\cdot)) = \prod_{i=1}^n \prod_{a=1}^k [q_{ai}(t_{ai})]^{\delta_{ai}} \frac{S_{ai}(t_{ai})}{S_{ai}(t_{ai}^*)} \quad (3.3)$$

em que $\mathbf{q}^{(0)}(\cdot) = (q_1^{(0)}(t), \dots, q_k^{(0)}(t))^\top$ representa o vetor das taxas iniciais, $q_{ai}(t_{ai})$ a função de risco ou taxa de transição, δ_{ai} é o indicador de falha (1, se t_{ai} é um tempo de falha na transição a para o indivíduo i ou 0, se t_{ai} é um tempo de censura na transição a para o indivíduo i), $S_{ai}(t_{ai})$ é a função de sobrevivência considerando o tempo de falha/censura e $S_{ai}(t_{ai}^*)$ o tempo de truncamento à esquerda.

A função de sobrevivência do indivíduo i na transição a é definida por

$$\begin{aligned} S_{ai}(t_{ai}) &= \exp\{-\exp\{(\boldsymbol{\beta}_j^\top \mathbf{x}_{ai})\} q_j^{(0)}(t_{ai})\} \\ S_{ji}(t_{ai}^*) &= \exp\{-\exp\{(\boldsymbol{\beta}_j^\top \mathbf{x}_{ai})\} q_a^{(0)}(t_{ai}^*)\}. \end{aligned} \quad (3.4)$$

Substituindo as equações em (3.4) na (3.3), tem-se a seguinte função de verossimilhança

$$L(\boldsymbol{\beta}, \mathbf{q}^{(0)}(\cdot)) = \prod_{i=1}^n \prod_{a=1}^k [q_{ai}^{(0)}(t_{ai}) \exp\{\boldsymbol{\beta}_a^\top \mathbf{x}_{ai}(t_{ai})\}]^{\delta_{ai}} \frac{\exp\{-\exp\{(\boldsymbol{\beta}_j^\top \mathbf{x}_{ai})\} q_j^{(0)}(t_{ai})\}}{\exp\{-\exp\{(\boldsymbol{\beta}_a^\top \mathbf{x}_{ai})\} q_a^{(0)}(t_{ai}^*)\}}$$

$$= \prod_{i=1}^n \prod_{a=1}^k [q_{ai}^{(0)}(t_{ai}) \exp\{\beta_a^\top \mathbf{x}_{ai}(t_{ai})\}]^{\delta_{ai}} \exp\{-\exp\{\beta_a^\top \mathbf{x}_{ai}\} \times [q_{ai}^{(0)}(t_{ai}) - q_{ai}^{(0)}(t_{ai}^*)]\}.$$

Por fim, o logaritmo da função de verossimilhança (log-verossimilhança) do modelo de Cox multiestados é definida por

$$l(\beta, \mathbf{q}^{(0)}(\cdot)) = \sum_{i=1}^n \sum_{a=1}^k \{\delta_{ai} \log[q_{ai}^{(0)}(t_{ai}) \exp(\beta_a^\top \mathbf{x}_{ji}(t_{ai}))] - \exp(\beta_a^\top \mathbf{x}_{ai}) [q_{ai}^{(0)}(t_{ai}) - q_{ai}^{(0)}(t_{ai}^*)]\}.$$

As estimações dos parâmetros são obtidas pela maximização da função de log-verossimilhança, em que é necessário recorrer a um processo iterativo como o de Newton-Raphson ou o Escore de Fisher.

3.3 Material e métodos

3.3.1 Material

Nessa seção, descrevem-se dois experimentos realizados na Escola Superior de Agricultura “Luiz de Queiroz” - USP, em que as variáveis de interesse são categóricas nominais e observadas ao longo do tempo (dados longitudinais). O objetivo é utilizar os dados para ajustar modelos de transição e consequentemente estimar as probabilidades das mudanças das categorias e suas respectivas taxas. A seguir, uma descrição desses dados.

3.3.1.1 Estudo sobre o comportamento do ácaro *Phytoseiulus persimilis*

O experimento foi conduzido no laboratório de patologia e controle microbiano de insetos por Canassa et al. (2019), com fêmeas do ácaro *Phytoseiulus persimilis*. Esse predador costuma localizar e atacar a praga *Tetranychus urticae* a fim de alimentar-se, e consequentemente combater a procriação da mesma. Dessa forma, tem-se interesse em estudar o comportamento do ácaro *Phytoseiulus persimilis* em plantas de feijão inoculadas com os fungos entomopatogênicos *Metarhizium robertsii* (extraído do solo de um campo de milho na cidade de Sinop, estado de Mato Grosso, Brasil) e *Beauveria bassiana* (extraído do solo de um campo de morango na cidade de Senador Amaral, estado de Minas Gerais, Brasil).

Neste ensaio, as fêmeas adultas *Tetranychus urticae* foram transferidas para dois tipos de discos foliares e os ácaros predadores privados de alimentos e posteriormente colocados no centro de uma ponte de *parafilm*, a partir da qual poderiam escolher entre os dois discos (com e sem tratamento fúngico). Para melhor desempenho, os dois discos foram aleatorizados e o clima da sala controlado, bem como o controle da iluminação. Foram preparadas dez arenas para cada tratamento fúngico. A Figura 3.1, ilustra como foi feita a separação dos dois discos foliares com relação à ponte de *parafilm*. Após os insetos serem introduzidos na ponte, observaram-se os possíveis comportamentos dos ácaros por vinte minutos qualificando-os nas categorias: 1 - procurando presa, 2 - encontrou presa, 3 - alimentando-se, 4 - andando fora da folha e 5 - andando na ponte de *parafilm*. Adicionalmente, o tempo, em minutos, para cada um desses estados foi mensurado, caracterizando um estudo longitudinal de tempo contínuo com variável resposta qualitativa nominal.

Figura 3.1: Esquema ilustrativo dos “caminhos de decisão” de fêmeas do ácaro *Phytoseiulus persimilis* no experimento conduzido por Canassa et al. (2019).

3.3.1.2 Estudo sobre o comportamento da microvespa *Telenomus podisi*

O experimento foi conduzido em Março de 2020 no Departamento de Entomologia (Laboratório de Biologia de Insetos), com fêmeas parasitoides de *Telenomus podisi*. Esse agente localiza os ovos do inseto hospedeiro *Euchistus heros* nos quais faz oviposição com a finalidade de reprodução da espécie. Assim sendo, o objetivo foi estudar o comportamento de forrageio das fêmeas parasitoides sobre ovos não parasitados e previamente parasitados dos percevejos. Consequentemente, aferir se as microvespas evitam a competição intraespecífica, ou seja, reconhecem ovos já parasitados e também se o tempo influencia a oviposição.

Neste experimento, as fêmeas parasitoides foram deixadas em um recipiente por 24 horas acasalando com os machos. Logo depois, foram transferidas para placas de petri que, aleatoriamente, continham quatro ovos não parasitados (sp0) ou previamente parasitados pela própria espécie (sp1), totalizando 16 repetições para cada tratamento. Destaque-se, também, que foram considerados o controle da luz e clima do ambiente. Em sequência, foram observados os possíveis comportamentos por dez minutos: 1 - voo ou caminhando, 2 - tamborilamento, 3 - oviposição ou 4 - marcação. Essas quatro categorias de comportamento são esquematizadas na Figura 3.2. O experimento descreve um estudo longitudinal de tempo contínuo com variável resposta qualitativa nominal.

Figura 3.2: Comportamentos da microvespa *Telenomus podisi*, associado ao experimento desenvolvido no Departamento de Entomologia e Acarologia da Escola Superior de Agricultura “Luiz de Queiroz” em Março de 2020, adaptado de Bin et al. (1993).

3.3.2 Métodos

A princípio, foi feita uma análise exploratória por meio de tabelas de contingências para os dois experimentos. No primeiro conjunto de dados, observaram-se cinco categorias de resposta, para descrever os comportamentos do ácaro predador *Phytoseiulus persimilis*. A distribuição inicial do processo pode ser observada na Figura 3.3a, com os possíveis caminhos a serem mensurados. No segundo conjunto, observaram-se quatro categorias de resposta, para os comportamentos da microvespa *Telenomus podisi*. Os possíveis caminhos podem ser observados na Figura 3.3b. Nos dois casos, ignora-se o fator tratamento.

(a) Comportamento de fêmeas do ácaro predador *Phytoseiulus persimilis*

(b) Comportamento de fêmeas parasitoides *Telenomus podisi*

Figura 3.3: Esquema do processo estocástico dos dois experimentos, com possíveis estados e probabilidades de transição a serem estimadas

Os comportamentos (estados) da microvespa ou ácaro predador foram classificadas como recorrentes ou não absorventes, isto é, dado que ocorreu um comportamento, a probabilidade de vir a ocorrer no futuro é diferente de zero.

Em sequência, ajustaram-se modelos multiestados, sob a hipótese de um processo *markoviano* com probabilidades e intensidades estacionárias no tempo, sem e com o efeito de tratamento nos dois experimentos. Em todos os modelos, considerou-se uma cadeia *markoviano* de alcance 1, ou seja, em que a categoria prévia é adicionalmente considerada no preditor linear, como é clássico nos modelos de multiestados. Para ajustar esses modelos, utilizou-se o método da máxima verossimilhança e considerou-se a distribuição exponencial para o período de tempo entre as transições conforme descrito na equação (3.1). Como os modelos em cada experimento são considerados encaixados, o efeito de tratamento foi verificado, por meio do teste de razão de verossimilhanças, cuja estatística é dada por

$$\Lambda = -2[\log(L(\hat{\theta}, \mathbf{x})) - \log(L(\tilde{\theta}, \mathbf{x}))], \quad (3.5)$$

$L(\hat{\theta}, \mathbf{x})$ e $L(\tilde{\theta}, \mathbf{x})$ correspondem às verossimilhanças dos modelos restrito (sem efeito de tratamento) e irrestrito (com efeito de tratamento), respectivamente. Rejeita-se a hipótese nula de que não há efeito de tratamento, se o valor da estatística (3.5) for maior do que o quantil da distribuição χ^2_ν , com $\nu = \dim(\tilde{\theta}) - \dim(\hat{\theta})$ graus de liberdade. Caso o teste de razão de verossimilhança leve à rejeição da hipótese nula, em cada experimento, as matrizes de probabilidades e intensidades do modelo multiestado serão

estimadas, para cada efeito de tratamento, permitindo a diferenciação entre seus níveis. Em termos práticos, o ajuste do modelo pelo método da máxima verossimilhança permite estimar as intensidades de cada comportamento e, pelo princípio da invariância, as probabilidades de transição são obtidas. A análise dos dados foi efetuada com auxílio do pacote **msm** (Jackson, 2018) disponível no *software R* (R Core Team, 2020).

3.4 Resultados e discussões

3.4.1 Análise dos dados sobre o ácaro *Phytoseiulus persimilis*

De início, foi realizada uma análise exploratória desconsiderando o efeito de tratamento fúngico. As frequências das transições da variável resposta comportamento do ácaro predador fêmea *Phytoseiulus persimilis* são apresentadas na Tabela 3.1. Nesse estudo, há um total de 178 transições, das quais 38 vêm do estado prévio 1 (procurando presa), 43 do estado prévio 2 (encontrou presa), 16 do estado prévio 3 (alimentando-se), 22 do estado prévio 4 (andando fora da folha) e 59 do estado prévio 5 (ponte de *parafilm*).

Tabela 3.1: Frequência observada e proporção (entre parênteses) das transições de comportamentos dos ácaros *Phytoseiulus persimilis*, em que 1: procurando presa, 2: encontrou presa, 3: alimentando-se, 4: andando fora da folha e 5: andando na ponte de *parafilm*.

Resposta prévia	Resposta futura					Total
	1	2	3	4	5	
1	23 (0,6053)	1 (0,0263)	12 (0,3158)	0 (0,0000)	2 (0,0526)	38
2	3 (0,0698)	23 (0,5349)	0 (0,0000)	15 (0,3488)	2 (0,0465)	43
3	4 (0,2500)	0 (0,0000)	12 (0,7500)	0 (0,0000)	0 (0,0000)	16
4	0 (0,0000)	7 (0,3182)	0 (0,0000)	15 (0,6818)	0 (0,0000)	22
5	14 (0,2373)	15 (0,2542)	0 (0,0000)	0 (0,0000)	30 (0,5085)	59
Total	44	46	24	30	34	178

Observou-se que houve menor frequência de transição para o estado em que o predador se encontrava “alimentando-se (3)”, e as maiores frequências estavam associadas ao estado “andando na ponte de *parafilm* (5)”. Com relação ao estado andando na ponte de *parafilm*, a probabilidade do ácaro permanecer nessa mesma categoria no futuro foi 0,5085, com 30 transições das 59 observadas. Contudo, essa análise é apenas exploratória e o fato de um ácaro permanecer em uma dada categoria não implica ausência de transição. Também associado ao seu comportamento está o tempo médio de permanência em cada estado. Essas características são melhores exploradas com o ajuste dos modelos.

Posteriormente, modelos multiestados restritos (sem efeito de tratamentos) e irrestritos (com efeito de tratamentos) foram ajustados aos dados. Com base no teste da razão de verossimilhanças, cuja estatística do teste foi $\Lambda = 33,8280$ associada a 17 graus de liberdade, foi verificada a existência do efeito dos tratamentos fúngicos nas plantas de feijão (p -valor = 0,0088). As estimativas das intensidades de transição para cada espécie são apresentadas na Tabela 3.2. Observe-se que, na diagonal de cada “matriz”, para cada tratamento, tem-se a intensidade de transição associada à permanência em cada estado. Quanto maior a taxa, menor é seu tempo médio de permanência. Por exemplo, para a espécie *Beauveria bassiana*, o tempo médio associado ao estado “alimentando-se” foi $\frac{-1}{-1,0275} \simeq 0,97$ segundos,

enquanto o tempo médio associado ao estado “andando sobre a ponte” foi $\frac{-1}{-0,0151} \simeq 66,22$ segundos. Com relação ao tratamento *Metarhizium robertsii*, o tempo médio referente ao estado “alimentando-se” foi $\frac{-1}{-0,4109} \simeq 2,44$ segundos, enquanto, para o estado “andando sobre a ponte”, foi $\frac{-1}{-0,5193} \simeq 1,93$ segundos. Registra-se, nesse sentido, que as menores intensidades de transição foram aquelas associadas ao tratamento *Beauveria bassiana* x *Metarhizium robertsii*, resultando em tempos médios de permanência maiores, quando comparados aos demais tratamentos. Com relação a mudança de comportamento “procurando presa” para “alimentando-se”, tem-se que os tratamentos *Beauveria bassiana* e *Metarhizium robertsii* retornaram maiores intensidades, 0,4272 e 0,2610 respectivamente. Enquanto a intensidade do tratamento *Beauveria bassiana* x *Metarhizium robertsii* foi 0,1082.

Tabela 3.2: Intensidades das transições de comportamento dos ácaros *Phytoseiulus persimilis*, em que 1: procurando presa, 2: encontrou presa, 3: alimentando-se, 4: andando fora da folha e 5: andando na ponte de *parafilm*, de acordo com cada tratamento fúngico.

Espécie	Resposta prévia	Resposta futura				
		1	2	3	4	5
<i>B. bassiana</i>	1	-0,4279	0,0000	0,4272	0,0000	0,0008
	2	0,0004	-0,4919	0,0000	0,4890	0,0025
	3	1,0275	0,0000	-1,0275	0,0000	0,0000
	4	0,0000	0,6344	0,0000	-0,6344	0,0000
	5	0,0053	0,0098	0,0000	0,0000	-0,0151
<i>M. robertsii</i>	1	-0,5820	0,0000	0,2610	0,0000	0,3210
	2	0,0000	-0,2503	0,0000	0,0123	0,2379
	3	0,4109	0,0000	-0,4109	0,0000	0,0000
	4	0,0000	0,0365	0,0000	-0,0365	0,0000
	5	0,3217	0,1976	0,0000	0,0000	-0,5193
<i>B. bassiana</i> x <i>M. robertsii</i>	1	-0,1223	0,0000	0,1082	0,0000	0,0141
	2	0,0006	-0,0559	0,0000	0,0522	0,0000
	3	0,1165	0,0000	-0,1165	0,0000	0,0000
	4	0,0000	0,0566	0,0000	-0,0566	0,0000
	5	0,0257	0,0043	0,0000	0,0000	-0,0300

Os tempos médios e respectivos desvios-padrão em cada estado para os tratamentos estão dispostos na Tabela 3.3. Esses resultados reforçam os cálculos feitos a partir das intensidades de permanência em cada estado dos tratamentos (Tabela 3.2), em que os tempos de permanência do tratamento *Beauveria bassiana* x *Metarhizium robertsii* foram os maiores em comparação com os demais.

Tabela 3.3: Média e desvio-padrão dos tempos de permanência em cada um dos comportamentos dos ácaros *Phytoseiulus persimilis*, em que 1: procurando presa, 2: encontrou presa, 3: alimentando-se, 4: andando fora da folha e 5: andando na ponte de *parafilm*, considerando-se cada tratamento fúngico.

Espécie	Medidas	Comportamentos				
		1	2	3	4	5
<i>B. bassiana</i>	Média	2,3367	2,0329	0,9732	1,5763	66,3140
	Desvio-padrão	3,9891	2,4092	1,7050	1,9584	34,9857
<i>M. robertsii</i>	Média	1,7183	3,9960	2,4337	27,3772	1,9258
	Desvio-padrão	2,5647	5,3935	5,5557	52,1173	1,9258
<i>B. bassiana</i> x <i>M. robertsii</i>	Média	8,1744	17,9028	8,5864	17,6780	33,3612
	Desvio-padrão	9,5522	16,2389	12,4854	18,2164	31,9101

Com a estimação das intensidades de transições, estimaram-se as probabilidades dessas mudanças, que são apresentadas na Figura 3.4. As cores mais fortes indicam as maiores probabilidades de transição. Com isso, constata-se visualmente uma diferença entre os três tratamentos, justificando o efeito significativo de tratamento detectado pelo teste da razão de verossimilhanças. Verificou-se que as maiores probabilidades de permanência no mesmo estado são para o composto *Beauveria bassiana* x *Metarhizium robertsii*, ou seja, sob esse tratamento fúngico, as probabilidades de mudanças de estados são pequenas, embora nenhum dos comportamentos seja um estado absorvente. Em todos os tratamentos, foram observados probabilidades não nulas e com relação as mudanças de estados para “alimentando-se” notou-se probabilidades baixas. Contudo, a mudança do estado 1 (procurando presa) para o estado 3 (alimentando-se) retornou maiores probabilidades para os tratamentos *Beauveria bassiana* e *Metarhizium robertsii*, respectivamente 0,225 e 0,165.

As probabilidades estimadas mostradas na Figura 3.4 têm relação direta com as taxas de intensidade de transição apresentadas na Tabela 3.2.

Figura 3.4: Estimativas das probabilidades do ácaro *Phytoseiulus persimilis*, em que 1: procurando presa, 2: encontrou presa, 3: alimentando-se, 4: andando fora da folha e 5: andando na ponte de parafilm, considerando-se cada tratamento fúngico.

Para uma melhor visualização dos comportamentos do ácaro *Phytoseiulus persimilis*, obtiveram-se as probabilidades das mudanças dos estados procurando presa ou andando fora da folha ou andando na ponte de *parafilm* para alimentando-se, considerando o passar do tempo em segundos (Figura 3.5). O tratamento *Metarhizium robertsii* indicou probabilidades constantes em torno de 0,20 em todas as mudanças de transição à medida que tempo passava em segundos. Já o composto *Beauveria bassiana* x *Metarhizium robertsii* retornou probabilidades maiores nas transações procurando presa ou andando na ponte de *parafilm* para alimentando-se. Entretanto, à medida que tempo passava essas probabilidades diminuem. Com relação ao tratamento *Beauveria bassiana*, observou-se probabilidades inferiores em comparação com os demais tratamentos. Nesse tratamento, na mudança procurando presa para alimentando-se, à medida que tempo passava as probabilidades diminuem mas aproxima-se das probabilidades estimadas do tratamento *Beauveria bassiana* x *Metarhizium robertsii*. Na mudança andando fora da folha para alimentando-se, à medida que tempo passava as probabilidades aumentavam e foram maiores que o do composto *Beauveria bassiana* x *Metarhizium robertsii*. Com relação a mudança andando na ponte de *parafilm* para alimentando-se, retornou as menores probabilidades. Contudo, à medida que tempo passava as probabilidades aumentavam-se e aproximavam-se dos demais tratamentos. Considerando a mudança de transição procurando presa para alimentando-se, tem-se que a razão de chance foi aproximadamente 3 vezes maior para o composto *Beauveria bassiana* x *Metarhizium robertsii* em relação ao tratamento *Metarhizium robertsii*, ao passarem 50 segundos. Já com relação ao tratamento *Beauveria bassiana*, o composto foi 1,6 vezes maior. Passando 200 segundos e considerando a mudança da ponte de *parafilm* para alimentando-se, retornou 3,6 vezes maior para o composto em relação ao tratamento *Beauveria bassiana* e 2 vezes maior do que o tratamento *Metarhizium robertsii*. Assim, as chances do ácaro *Phytoseiulus persimilis* alimentar-se foram maiores quando submetido ao tratamento fúngico *Beauveria bassiana* x *Metarhizium robertsii*.

Figura 3.5: Estimativas das probabilidades do ácaro *Phytoseiulus persimilis* do estado 1 para 3, 4 para 3 e 5 para 3, considerando cada tratamento (Bb - *Beauveria bassiana*, Mr - *Metarhizium robertsii* e BM - *Beauveria bassiana* x *Metarhizium robertsii*) e a mudança do tempo.

3.4.2 Análise dos dados sobre o parasitoide *Telenomus podisi*

A análise desse segundo experimento segue as mesmas etapas da análise anterior. As frequências das transições, da variável resposta comportamento de fêmeas parasitóides *Telenomus podisi* são apresentadas na Tabela 3.4. Observa-se um total de 128 transições, das quais 54 vêm do estado prévio 1 (voo ou caminhando), 23 do estado prévio 2 (tamborilando), 45 do estado prévio 3 (oviposição) e 6 do estado prévio 4 (marcando). Com relação ao estado prévio, nota-se que houve menor frequência de transição no comportamento “marcando” e maior frequência na categoria “voo caminhando”. Além disso, as probabilidades da microvespa permanecer nos estados “voo caminhando” (0,61) e “oviposição” (0,64) foram maiores quando comparadas as dos estados “tamborilando” (0,39) e “marcando” (0,00). Contudo, essa análise é apenas exploratória e o fato da *Telenomus podisi* permanecer em um dada categoria não implica em ausência de transição.

Tabela 3.4: Frequência observada e proporção (entre parênteses) das microvespas parasitoides *Telenomus podisi*, em que 1: voo ou caminhando, 2: tamborilando, 3: oviposição e 4: marcando

Resposta prévia	Resposta futura				Total
	1	2	3	4	
1	33 (0,6111)	13 (0,2407)	8 (0,1481)	0 (0,0000)	54
2	4 (0,1739)	9 (0,3913)	10 (0,4348)	0 (0,0000)	23
3	1 (0,0222)	8 (0,1777)	29 (0,6444)	7 (0,1555)	45
4	0 (0,0000)	0 (0,0000)	6 (1,0000)	0 (0,0000)	6
Total	38	30	53	7	128

Os modelos multiestados restritos (sem efeito de tratamentos) e irrestritos (com efeito de tratamentos) foram ajustados aos dados para estimar o tempo médio de permanência, nos comportamentos da microvespa parasitoide *Telenomus podisi*. Com base no teste da razão de verossimilhanças, cuja estatística foi $\Lambda = 37,6790$ associada a 11 graus de liberdade, verificou-se a existência do efeito dos tratamentos (sp0 e sp1) (p -valor = 0,0001). As estimativas das intensidades de transições dos comportamentos para cada tratamento são apresentadas na Tabela 3.5. Na diagonal de cada “matriz”, tem-se a intensidade de transição associada à permanência em cada estado. Quanto maior a taxa, menor é seu tempo médio de permanência no estado. Por exemplo, para os ovos não parasitados (sp0), o tempo médio associado ao estado “marcando” foi, aproximadamente, 31 segundos ($\frac{-1}{-0,0322}$), enquanto o tempo médio associado ao previamente parasitado (sp1) foi aproximadamente 1 segundo ($\frac{-1}{-1,3367}$). Vale ressaltar que uma intensidade de transição nula não implica ausência de intensidades em outros comportamentos.

Tabela 3.5: Intensidades das transições do comportamento da microvespa *Telenomus podisi*, em que 1: voo ou caminhando, 2: tamborilando, 3: oviposição e 4: marcando, de acordo com cada tratamento (sp0 - ovos não parasitados e sp1 - ovos previamente parasitados).

Tratamento	Resposta prévia	Resposta futura			
		1	2	3	4
sp0	1	-0,0075	0,0023	0,0052	0,0000
	2	0,0022	-0,0120	0,0097	0,0001
	3	0,0000	0,0017	-0,0074	0,0057
	4	0,0000	0,0001	0,0321	-0,0322
sp1	1	-0,0037	0,0037	0,0000	0,0000
	2	0,0029	-0,0091	0,0060	0,0002
	3	0,0000	0,0106	-0,0115	0,0010
	4	0,0000	0,0001	1,3366	-1,3367

As estimativas dos tempos médios e respectivos desvios-padrão, para cada comportamento, estão dispostos na Tabela 3.6. Observou-se que as permanências para os comportamentos oviposição e marcação são maiores quando as microvespas tinham disponíveis ovos não parasitados (sp0). Já para os ovos previamente parasitados (sp1), os tempos médios foram maiores nos estados voo caminhando e tamborilando. Ressalta-se que na categoria oviposição o desvio-padrão foi menor para o tratamento ovos não parasitados (sp0).

Tabela 3.6: Média e desvio-padrão dos tempos de permanência em cada um dos comportamentos da microvespa *Telenomus podisi*, em que 1: voo ou caminhando, 2: tamborilando, 3: oviposição e 4: marcando, considerando-se cada tratamento (sp0 e sp1).

Tratamento	Medidas	Comportamentos			
		1	2	3	4
sp0	Média	134,1266	82,9912	134,4171	31,0951
	Desvio-padrão	48,1854	44,8590	101,2770	30,2004
sp1	Média	271,8204	110,1277	86,6757	0,7481
	Desvio-padrão	86,5264	49,5328	530,8702	53,6810

Com a estimação das intensidades de transições obtiveram-se as probabilidades, que são apresentadas na Figura 3.6. As cores mais fortes indicam maiores probabilidades de transição, constatando-se visualmente que as maiores probabilidades referem-se às permanências no mesmo estado para os dois tratamentos. Destaca-se ainda que no tratamento sp1 (ovos previamente parasitados), a probabilidade de transição da categoria “marcando” para “oviposição” foi 0,7323.

Figura 3.6: Estimativas das probabilidades da microvespa *Telenomus podisi*, em que 1: voo ou caminhando, 2: tamborilando, 3: oviposição e 4: marcando, considerando-se cada tratamento

A fim de ter uma melhor visualização desses comportamentos, obtiveram-se as probabilidades das mudanças da microvespa *Telenomus podisi* considerando o passar do tempo em segundos (Figura 3.7). Verificou-se que, em todas transições, as probabilidades de oviposição das microvespas foram maiores para ovos não parasitados. Ao calcular a razão de chances, considerando a mudança voo caminhando para oviposição, obteve-se um valor aproximadamente de 13 vezes maior, quando submetidos a ovos não parasitados, ao passarem 50 segundos. Ao passar 200 segundos e considerando a mudança do comportamento tamborilando para oviposição, obteve-se um valor aproximadamente 5 vezes maior quando expostos a ovos não parasitados, indicando assim maiores chances, quando a microvespa *Telenomus podisi* é submetida ao tratamento sp0 (ovos não parasitados).

Figura 3.7: Estimativas das probabilidades da microvespa *Telenomus podisi* do estado 1 para 3, 2 para 3 e 4 para 3, considerando cada tratamento (sp0 - ovos não parasitados e sp1 - ovos previamente parasitados) e a mudança do tempo.

3.5 Conclusão

Os modelos multiestados são capazes de estudar variáveis dependentes do tipo categóricas longitudinais. Nesses casos é possível obter as matrizes de transição e intensidade, as quais resumem sobremaneira o estudo de tempo contínuo.

Nesse trabalho, a utilização dessa metodologia possibilitou a análise de duas situações típicas de estudos entomológicos, em que havia interesse na descrição do comportamento do ácaro *Phytoseiulus persimilis* e da microvespa *Telenomus podisi* sob efeito de tratamentos. Não se observaram, nesses estudos, dados de censura. Porém, mesmo sob essa restrição, a metodologia foi versátil.

No experimento sobre as fêmeas do ácaro predador *Phytoseiulus persimilis*, os resultados sinalizam uma maior atividade desse indivíduo quando submetidos aos tratamentos *Beauveria bassiana* e *Metarhizium robertsii*. Entretanto, ao calcular a razão de chances dos tratamentos no que diz respeito à mudança do estado “procurando presa” ou “andando na ponte de *parafilm*” para “alimentando-se”, o resultado indicou um maior favorecimento ao tratamento composto *Beauveria bassiana* x *Metarhizium robertsii*. Portanto, o ácaro predador *Phytoseiulus persimilis* é mais propício a alimentar-se das fêmeas adultas *Tetranychus urticae* quando submetido ao tratamento composto. No segundo experimento sobre a microvespa *Telenomus podisi*, os resultados sinalizam maior desempenho e chances de oviposição da fêmea, quando submetidas ao tratamento sp0 (ovos não parasitados), indicando que as microvespas *Telenomus podisi* evitam competição intraespecífica.

Referências

- Behie, S. W., Jones, S. J., e Bidochka, M. J. (2015). Plant tissue localization of the endophytic insect pathogenic fungi metarhizium and beauveria. *Fungal Ecology*, 13:112–119.
- Bin, F., Vinson, S., Strand, M., Colazza, S., e Jones Jr, W. (1993). Source of an egg kairomone for trissolcus basalis, a parasitoid of nezara viridula. *Physiological Entomology*, 18(1):7–15.
- Canassa, F., Tall, S., Moral, R. A., de Lara, I. A., Delalibera Jr, I., e Meyling, N. V. (2019). Effects of bean seed treatment by the entomopathogenic fungi metarhizium robertsii and beauveria bassiana on plant growth, spider mite populations and behavior of predatory mites. *Biological Control*, 132:199–208.
- Conn, P. B. e Cooch, E. G. (2009). Multistate capture–recapture analysis under imperfect state observation: an application to disease models. *Journal of Applied Ecology*, 46(2):486–492.
- Gámez-Virués, S., Gurr, G., Raman, A., La Salle, J., e Nicol, H. (2009). Effects of flowering groundcover vegetation on diversity and activity of wasps in a farm shelterbelt in temperate australia. *BioControl*, 54(2):211–218.
- Gimenez, O., Choquet, R., e Lebreton, J.-D. (2003). Parameter redundancy in multistate capture–recapture models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 45(6):704–722.
- Hoel, P. G., Port, S. C., e Stone, C. J. (1972). *Introduction to Stochastic Processes*. Houghton Mifflin Company, Boston.
- Jaber, L. R. e Enkerli, J. (2016). Effect of seed treatment duration on growth and colonization of vicia faba by endophytic beauveria bassiana and metarhizium brunneum. *Biological Control*, 103:187–195.
- Jackson, C. (2018). Multi-state modelling with r: the msm package. *Cambridge, UK*.

- Kalbfleisch, J. e Lawless, J. F. (1985). The analysis of panel data under a markov assumption. *Journal of the American Statistical Association*, 80(392):863–871.
- Karlin, S. e Taylor, H. (1975). *A first course in stochastic processes*. Academic Press, New York.
- Klein, J. P. e Moeschberger, M. L. (2006). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.
- Lara, I., Hinde, J., de Castro, A., e da Silva, I. (2017). A proportional odds transition model for ordinal responses with an application to pig behaviour. *Journal of Applied Statistics*, 44(6):1031–1046.
- Lara, I. A. R., Hinde, J., e Taconeli, C. A. (2018). An alternative method for evaluating stationarity in transition models. *Journal of Statistical Computation and Simulation*, 87(15):2962–2980.
- Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suárez, C., e Andersen, P. K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical methods in medical research*, 18(2):195–222.
- Molenberghs, G. e Verbeke, G. (2006). *Models for discrete longitudinal data*. Springer Science & Business Media.
- Oliveira, R. d. V. C. d., Shimakura, S. E., Campos, D. P., Hökerberg, Y. H. M., Victoriano, F. P., Ribeiro, S., Veloso, V. G., Grinsztejn, B., e Carvalho, M. S. (2018). Effects of antiretroviral treatment and nadir cd4 count in progression to cardiovascular events and related comorbidities in a hiv brazilian cohort: a multi-stage approach. *AIDS care*, 30(5):551–559.
- Oliveira, R. d. V. C. d., Shimakura, S. E., Campos, D. P., Victoriano, F. P., Ribeiro, S. R., Veloso, V. G., Grinsztejn, B., e Carvalho, M. S. (2013). Multi-state models for defining degrees of chronicity related to hiv-infected patient therapy adherence. *Cadernos de saude publica*, 29:801–811.
- Prezoto, F., Maciel, T. T., Detoni, M., Mayorquin, A. Z., e Barbosa, B. C. (2019). Pest control potential of social wasps in small farms and urban gardens. *Insects*, 10(7):192.
- Putter, H. (2011). Tutorial in biostatistics: Competing risks and multi-state models analyses using the mstate package. *Companion file for the mstate package*.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rotolo, F. (2013). Frailty multi-state models for the analysis of survival data from multicenter clinical trials.
- Singer, B. e Spilerman, S. (1976). The representation of social processes by markov models. *American journal of sociology*, 82(1):1–54.
- Wang, Z.-z., Liu, Y.-q., Min, S., Huang, J.-h., e Chen, X.-x. (2019). Parasitoid wasps as effective biological control agents. *Journal of Integrative Agriculture*, 18(4):705–715.
- Ware, J. H., Lipsitz, S., e Speizer, F. E. (1988). Issues in the analysis of repeated categorical outcomes. *Statistics in Medicine*, 7(1-2):95–107.
- Wreede, L. C., Fiocco, M., e Putter, H. (2010). The mstate package for estimation and prediction in non-and semi-parametric multi-state and competing risks models. *Computer methods and programs in biomedicine*, 99(3):261–274.
- Wreede, L. C., Fiocco, M., Putter, H., et al. (2011). mstate: an r package for the analysis of competing risks and multi-state models. *Journal of statistical software*, 38(7):1–30.

4 CONSIDERAÇÕES FINAIS

O presente trabalho apresenta uma alternativa para analisar experimentos entomológicos sobre controle biológico de pragas em estudos com comportamentos de insetos e ácaros, desde que esses ensaios resultem em dados categóricos nominais com medidas repetidas ao longo do tempo. Assim, os modelos de transição podem ser utilizados para auxiliar as tomadas de decisões na área.

Os modelos de transição de tempo contínuo são mais informativos do que os modelos discretos. A mensuração do tempo permite incorporar parâmetros associados ao tempo do processo. Independente do processo ser contínuo ou discreto a suposição de estacionariedade em modelos de transição é sempre desejável, mas nem sempre ocorre devido a variações aleatórias. Quando existe a violação dessa suposição, tem-se um processo não homogêneo. Consequentemente, um modelo de transição mais complexo. Visto que, é necessário estimar t matrizes de transição. Dessa forma, é importante testar essa condição para a escolha do modelo correto. Nesse trabalho, consideram-se testes para o caso discreto, mas no processo contínuo essa verificação também é necessária.

A partir do estudo de simulação realizado no capítulo 2, a estatística gradiente obteve bons resultados. Mostrando-se ser competitiva com as estatísticas de Anderson e Goodman, *Wald* e razão de verossimilhança. Com o auxílio desse teste assume-se estacionariedade para a análise dos dados de microvespa *Telenomus podisi*, constatando-se por meio do modelo que há uma preferência a ovos não parasitados de *Euschistus heros*. Por esse motivo, a microvespa *Telenomus podisi* impossibilita a competição dentro da própria espécie.

No capítulo 3, consideraram-se os modelos multiestados associados aos experimentos de tempo contínuo. No ensaio sobre ácaro predador *Phytoseiulus persimilis*, verificou-se uma maior predação de presas que se alimentam de plantas tratadas pelo composto *Beauveria bassiana* x *Metarhizium robertsii*, com base nas razões de chances das mudanças para alimentar-se da presa obtidas em relação aos demais tratamentos.

A análise sob a ótica de tempo contínuo do estudo da microvespa *Telenomus podisi*, ratificou sua preferência em ovipositar em ovos não parasitados de *Euschistus heros*. Ressalta-se a necessidade de desenvolver métodos para avaliar a estacionariedade ou não desses processos contínuos, pois este é um fator que pode mudar as conclusões. Uma vez que, utilizar apenas uma matriz de transição (processo estacionário) para interpretar os resultados de t matrizes (processo não estacionário) levaria a interpretações errôneas.

Como sugestão para trabalhos futuros, pode-se estudar análise de resíduos e diagnósticos, bem como a avaliação da qualidade do ajuste nesses modelos.

5 APÊNDICES

5.1 Distribuição Multinomial

$$X \sim \text{Mult.}(n, \pi_1, \pi_2, \dots, \pi_k)$$

$$P(X_1 = n_1, \dots, X_k = n_k) = \frac{n!}{n_1! \dots n_k!} \pi_1^{n_1} \pi_2^{n_2} \dots \pi_k^{n_k}$$

Função de verossimilhança

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n \prod_{j=1}^J \prod_{t=1}^T [P(X_1 = n_1, \dots, X_k = n_k)] \\ &= \prod_{i=1}^n \prod_{j=1}^J \prod_{t=1}^T [\pi_{ab}(t)(\mathbf{x})]^{y_{ijt}} \end{aligned}$$

em que π_{ab} é a probabilidade marginal e y_{ijt} é uma função indicadora.

Logaritmo da função de verossimilhança

$$\ell(\theta) = \log L(\theta) = \log \prod_{i=1}^n \prod_{j=1}^J \prod_{t=1}^T [\pi_{ab}(t)(\mathbf{x})]^{y_{ijt}} = \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^T \log \{ [\pi_{ab}(t)(\mathbf{x})]^{y_{ijt}} \}$$

5.1.1 Logaritmo da função de verossimilhança para modelo logito generalizado de transição

$$\begin{aligned} \ell(\theta) &= \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^T \log \{ [\pi_{ab}(t)(\mathbf{x})]^{y_{ijt}} \} = \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^T \{ y_{ijt} \log [\pi_{ab}(t)(\mathbf{x})] \} \\ &= \sum_{i=1}^n \sum_{t=1}^T \left\{ \sum_{j=1}^{J-1} y_{ijt} \log \left[\frac{\exp(\lambda_{ab}(t) + \delta_t^T \mathbf{x})}{1 + \sum_{j=1}^{J-1} \exp(\lambda_{ab}(t) + \delta_t^T \mathbf{x})} \right] \right\} \\ &= \sum_{i=1}^n \sum_{t=1}^T \left\{ \sum_{j=1}^{J-1} y_{ijt} \log \{ \exp(\lambda_{ab}(t) + \delta_t^T \mathbf{x}) \} - \log \{ 1 + \sum_{j=1}^{J-1} \exp(\lambda_{ab}(t) + \delta_t^T \mathbf{x}) \} \right\} \\ &= \sum_{i=1}^n \sum_{t=1}^T \left\{ \sum_{j=1}^{J-1} y_{ijt} (\lambda_{ab}(t) + \delta_t^T \mathbf{x}) - \log \{ 1 + \sum_{j=1}^{J-1} \exp(\lambda_{ab}(t) + \delta_t^T \mathbf{x}) \} \right\} \end{aligned}$$

1ª Derivada

$$\begin{aligned} \mathbf{U}(\lambda_{ab}) &= \sum_{i=1}^n \sum_{t=1}^T \left\{ \sum_{j=1}^{J-1} y_{ijt} - \frac{\sum_{j=1}^{J-1} \exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})}{1 + \sum_{j=1}^{J-1} \exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})} \right\} \\ \mathbf{U}(\delta_t) &= \sum_{i=1}^n \sum_{t=1}^T \left\{ \sum_{j=1}^{J-1} y_{ijt} \mathbf{x} - \frac{\sum_{j=1}^{J-1} \exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x}) \mathbf{x}}{1 + \sum_{j=1}^{J-1} \exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})} \right\} \end{aligned}$$

5.1.2 Logaritmo da função de verossimilhança para modelo chances proporcionais de transição

$$\begin{aligned} \ell(\theta) &= \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^T \log \{ [\pi_{ab}(t)(\mathbf{x})]^{y_{ijt}} \} = \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^T \{ y_{ijt} \log [\pi_{ab}(t)(\mathbf{x})] \} \\ &= \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^T \{ y_{ijt} \log [\gamma_{ab}(t)(\mathbf{x}) - \gamma_{a(b-1)}(t)(\mathbf{x})] \} \\ &= \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^T \left\{ y_{ijt} \log \left[\frac{\exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})}{1 + \exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})} - \frac{\exp(\lambda_{a(b-1)}(t) + \delta_t^T \mathbf{x})}{1 + \exp(\lambda_{a(b-1)}(t) + \delta_t^T \mathbf{x})} \right] \right\} \\ &= \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^T \left\{ y_{ijt} \log \left[\frac{A - B}{[1 + \exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})][1 + \exp(\lambda_{a(b-1)}(t) + \delta_t^T \mathbf{x})]} \right] \right\} \end{aligned}$$

em que

$$\begin{aligned} A &= \exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x}) [1 + \exp(\lambda_{a(b-1)}(t) + \delta_t^T \mathbf{x})] \\ &= \exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x}) + \exp(\lambda_{ab(t)} + \lambda_{a(b-1)}(t) + 2\delta_t^T \mathbf{x}) \end{aligned}$$

$$\begin{aligned} B &= \exp(\lambda_{a(b-1)}(t) + \delta_t^T \mathbf{x}) [1 + \exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})] \\ &= \exp(\lambda_{a(b-1)}(t) + \delta_t^T \mathbf{x}) + \exp(\lambda_{ab(t)} + \lambda_{a(b-1)}(t) + 2\delta_t^T \mathbf{x}) \end{aligned}$$

$$\begin{aligned} \ell(\theta) &= \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^T \left\{ y_{ijt} \log \left[\frac{\exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x}) - \exp(\lambda_{a(b-1)}(t) + \delta_t^T \mathbf{x})}{[1 + \exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})][1 + \exp(\lambda_{a(b-1)}(t) + \delta_t^T \mathbf{x})]} \right] \right\} \\ &= \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^T y_{ijt} \{ \log [\exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x}) - \exp(\lambda_{a(b-1)}(t) + \delta_t^T \mathbf{x})] - \\ &\quad - \log [1 + \exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})] - \log [1 + \exp(\lambda_{a(b-1)}(t) + \delta_t^T \mathbf{x})] \} \end{aligned}$$

1ª Derivada

$$\mathbf{U}(\lambda_{ab}) = \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^T y_{ijt} \left\{ \frac{\exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})}{\exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x}) - \exp(\lambda_{a(b-1)(t)} + \delta_t^T \mathbf{x})} - \frac{\exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})}{1 + \exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})} \right\}$$

$$\begin{aligned} \mathbf{U}(\delta_t) &= \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^T y_{ijt} \left\{ \frac{\exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x}) \mathbf{x} - \exp(\lambda_{a(b-1)(t)} + \delta_t^T \mathbf{x}) \mathbf{x}}{\exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x}) - \exp(\lambda_{a(b-1)(t)} + \delta_t^T \mathbf{x})} - \right. \\ &\quad \left. - \frac{\exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x}) \mathbf{x}}{1 + \exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})} - \frac{\exp(\lambda_{a(b-1)(t)} + \delta_t^T \mathbf{x}) \mathbf{x}}{1 + \exp(\lambda_{a(b-1)(t)} + \delta_t^T \mathbf{x})} \right\} \\ &= \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^T y_{ijt} \mathbf{x} \left\{ 1 - \frac{\exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})}{1 + \exp(\lambda_{ab(t)} + \delta_t^T \mathbf{x})} - \frac{\exp(\lambda_{a(b-1)(t)} + \delta_t^T \mathbf{x})}{1 + \exp(\lambda_{a(b-1)(t)} + \delta_t^T \mathbf{x})} \right\} \end{aligned}$$