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

Métodos estatísticos aplicados ao teste de *Salmonella*/microsoma:
modelos, seleção e suas implicações

Davi Butturi-Gomes

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

Piracicaba
2016

Davi Butturi-Gomes
Ecólogo

**Métodos estatísticos aplicados ao teste de *Salmonella*/microsoma: modelos,
seleção e suas implicações**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **SILVIO SANDOVAL ZOCCHI**

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

Piracicaba
2016

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

Butturi-Gomes, Davi

Métodos estatísticos aplicados ao teste de *Salmonella*/microssoma: modelos, seleção e suas implicações / Davi Butturi-Gomes. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2016.
180 p. : il.

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

1. Modelos não-lineares generalizados 2. Modelos lineares generalizados vetoriais
3. Distribuição de Skellam 4. Teste de Ames 5. Funções em R I. Título

CDD 589.95
B989m

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

*Com carinho, dedico este trabalho aos meus
sobrinhos Natália, Leonardo e Francisco.*

AGRADECIMENTOS

Ao Prof. Dr. Silvio Sandoval Zocchi, pela orientação e apoio.

Aos professores Dra. Clarice Borges Demétrio, Dr. Cristian M. V. Lobos, Dra. Cecília Candolo, Dra. Roseli Ap. Leandro, Dra. Maria Aparecida Marin-Morales, Dr. Miguel Petrere Jr., Dr. Thomas W. Yee e Dr. David Anderson pelos ensinamentos e pela colaboração no desenvolvimento deste trabalho.

À Dra. Cintya Christofolletti, à MSc. Yaliana Tafurt e, principalmente, à Dra. Thaís C. C. Fernandes pela parceria e pelos ensinamentos, mas também pelas risadas em dias cansativos de laboratório, e pela amizade.

Aos colegas de pós-graduação Iuri, Natália, Everton, Luiz Ricardo, Maria, Ricardo, José Nilton, Thiago Gentil, Lucas, Alessandra, Gilson, Elisângela; com quem tive o prazer de conviver.

Aos funcionários do Departamento de Ciências Exatas (LCE-ESALQ/USP) e do Laboratório de Mutagênese (LabMut-IB/UNESP Rio Claro).

À entidade financiadora, que possibilitou o desenvolvimento deste trabalho em dedicação integral, Fundação CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

Aos amigos de sempre, Jaca e Elen.

Aos meus irmãos, cunhados e à família, de Luiz e Ana Neusa, que me acolheu.

Aos meus pais Rosa e Milton.

Ao Nando.

No alheamento da obscura forma humana,
De que, pensando, me desencarcerou,
Foi que eu, num grito de emoção, sincero,
Encontrei, afinal, o meu Nirvana.

Nessa manumissão schopenhaueriana,
Onde a Vida do humano aspecto fero
Se desarraiga, eu, feito força, impero
Na imanência da Idéia Soberana!

Destruída a sensação que oriunda fora
Do tato – ínfima antena aferidora
Destas tegumentárias mãos plebéias –

Gozo o prazer, que os anos não carcomem,
De haver trocado a minha forma de homem
Pela imortalidade das Idéias!

— *Augusto dos Anjos,*
“O Meu Nirvana”.

SUMÁRIO

RESUMO.....	11
ABSTRACT.....	13
LISTA DE FIGURAS.....	15
LISTA DE TABELAS	19
1 INTRODUÇÃO.....	23
2 REVISÃO BIBLIOGRÁFICA	27
2.1 Teste de <i>Salmonella</i> /microsoma	27
2.2 Modelos propostos para o teste de <i>Salmonella</i> /microsoma	30
2.3 Modelos não-lineares generalizados	34
2.3.1 Suposições e fórmula geral	34
2.3.2 Estimação	36
2.3.3 Resíduos, diagnósticos e gráficos.....	38
2.3.4 Superdispersão	41
2.4 Modelos lineares generalizados vetoriais.....	46
2.4.1 Suposições e fórmula geral	46
2.4.2 Estimação	47
2.4.3 Resíduos, diagnósticos e gráficos.....	47
2.5 Seleção de modelos.....	49
2.5.1 Teste de Wald	49
2.5.2 Teste da razão de verossimilhanças	49
2.5.3 Critérios de Informação	50
3 MATERIAL.....	55
4 METODOLOGIA.....	59
4.1 Modelos tradicionais.....	59
4.1.1 Estudos de simulação	59
4.1.2 Obtenção dos valores iniciais.....	61
4.1.3 Ajuste dos modelos tradicionais.....	65
4.2 Modelos da família Skellam	66
4.3 Modelos não-lineares vetoriais da família binomial negativa.....	67
4.4 Estratégias para seleção de modelos, inferência e medidas de interesse.....	69
4.5 Implementações computacionais: o pacote “ames”	70
4.5.1 Função <code>fit.ames()</code>	71
4.5.2 Função <code>iv.ames()</code>	73
4.5.3 Função <code>dtest.ames()</code>	73
4.5.4 Função <code>plot.fitames()</code>	74
4.5.5 Função <code>diagnostics.fitames()</code>	75
4.5.6 Função <code>aic.fitames()</code>	76

4.5.7 Função <code>mmd.fitames()</code>	77
4.5.8 Função <code>rames()</code>	77
4.5.9 Funções de apoio	78
5 RESULTADOS E DISCUSSÃO	81
5.1 Modelos tradicionais.....	81
5.1.1 Estudos de simulação e problema do valor inicial	81
5.1.2 Ajuste dos modelos	101
5.2 Modelos propostos.....	145
5.3 Implicações dos modelos finais.....	155
6 CONCLUSÕES	163
REFERÊNCIAS	165
APÊNDICES.....	175

RESUMO

Métodos estatísticos aplicados ao teste de *Salmonella*/microsoma: modelos, seleção e suas implicações

O teste de *Salmonella*/microsoma é um ensaio biológico amplamente utilizado para avaliar o potencial mutagênico de substâncias que podem colocar em risco a saúde humana e a qualidade ambiental. A variável resposta é constituída pela contagem do número de colônias revertentes em cada placa, entretanto geralmente há dois efeitos confundidos, o de toxicidade e o de mutagenicidade. Alguns modelos foram propostos para a análise dos dados desses experimentos, que nem sempre apresentam bons ajustes e não consideram explicitamente interações. Há, ainda, poucas plataformas computacionais disponíveis que integram todas essas propostas e forneçam critérios para a seleção adequada de um modelo. Além disso, geralmente é difícil comparar os efeitos de diferentes substâncias sobre as várias linhagens da bactéria, então medidas com interpretação biológica direta são necessárias. Neste trabalho, foram investigadas as propriedades dos preditores dos modelos tradicionais, bem como o comportamento das distribuições amostrais dos estimadores dos parâmetros desses modelos, na presença de diversos níveis de superdispersão. Também, foram realizados experimentos com as linhagens TA98 e TA100 da bactéria, expostas aos inseticidas, metabolizados e não-metabolizados, Fipronil e Tiametoxam, dois agroquímicos bastante utilizados no Brasil. Aos dados desses experimentos foram ajustados diversos modelos, tanto aqueles tradicionalmente utilizados, quanto novos modelos, alguns baseados na regressão de Skellam e outros com interações explícitas. Para tal, foi obtida uma nova classe de modelos chamada de modelos não-lineares vetoriais generalizados e foi desenvolvido um pacote computacional em linguagem R, intitulado “ames”, para o ajuste, diagnóstico e seleção de modelos. Por fim, foram propostas medidas de interesse biológico, baseadas nos modelos selecionados, para avaliação de risco e do comprometimento do material genético e intervalos de confiança *bootstrap* paramétrico foram obtidos. Dentre os modelos tradicionais, aqueles cujas distribuições amostrais dos estimadores possuem melhor aproximação normal foram os de Bernstein, Breslow e Myers. Estes resultados forneceram um critério prático para a seleção de modelos, particularmente nas situações em que as medidas de AIC e de bondade de ajuste, os testes de razão de verossimilhanças e a análise de resíduos ou são pouco informativos ou simplesmente não podem ser aplicados. A partir dos modelos selecionados, pode-se concluir que a interação do fator de metabolização é significativa para a linhagem TA98 exposta ao Fipronil, tanto com relação aos efeitos tóxicos quanto aos efeitos mutagênicos; que o mecanismo de ação do Tiametoxam sobre a linhagem TA98 é completamente diferente quando o produto está metabolizado; e que, para a linhagem TA100, não houve efeito de metabolização considerando ambos os agroquímicos. Baseando-se nas medidas propostas, pode-se concluir que o Tiametoxam oferece os maiores riscos de contaminação residual, ainda que o Fipronil apresente os maiores índices de mutagenicidade.

Palavras-chave: Modelos não-lineares generalizados; Modelos lineares generalizados vetoriais; Distribuição de Skellam; Teste de Ames

ABSTRACT

Statistical methods applied for *Salmonella*/microsome test data: models, selection and their entailments

The *Salmonella*/microsome test is a widely accepted biological assay used to evaluate the mutagenic potential of substances, which can compromise human health and environment quality. The response variable in such experiments is typically the total number of reverts per plate, which, in turn, is the result of the confounded effects of mutagenicity and toxicity. Despite of some statistical models have already been established in the literature, they do not always fit well and neither explicitly consider interaction terms. Besides, there is just a number of available software able to handle these different approaches, usually lacking of global performance and model selection criteria. Also, it is often a hard task to compare the effects of different chemicals over the several available strains to perform the assay, and, thus, direct measures of biological implications are required. In this work, the properties of the predictors in each traditional model were investigated, as well as the behavior of the sampling distributions of the parameter estimators of these models, in different levels of overdispersion. Also, experiments using TA98 and TA100 strains were performed, by exposition to two insecticides, namely Fipronil and Thiamethoxam, currently used in Brazil, each of them prior and after to a metabolization processes. Then, the traditional models, empirical regression models based on the Skellam distribution and also compound mechanistic-empirical models with explicit interaction terms were fitted to the data. In order to use a single fitting framework, a new class of models was presented, namely the vector generalized nonlinear models, and a R language package, entitled “ames”, was developed for fitting, diagnosing and selection of models. Finally, some measures of biological interest were approached based on the selected models for the data, in the contexts of risk evaluation and of DNA damage cautioning. Confidence intervals for such measures were provided using bootstrap percentiles. Among the traditional models, the ones from Bernstein, Breslow and Myers were those whose sampling distributions presented the best normal approximations. These results provided a practical criterion for model selection, particularly in situations where measures as AIC and goodness of fit, likelihood ratio tests, and residual analysis are non informative or simply cannot be applied. From the final selected models, it was inferred that the interactions between the metabolization factor is significative for TA98 strain exposed to Fipronil, regarding both, mutagenic and toxic effects; that the dynamics between mutagenicity and toxicity are different when Thiamethoxam is metabolized compared to when it is not; and that there was no evidence to consider metabolization factor interactions for the TA100 strain data exposed to neither of the insecticides. By applying the referred measures of biological interest, it was concluded that the use of Thiamethoxam provides greater residual contamination risks and that Fipronil causes higher mutagenicity indices.

Keywords: Generalized nonlinear models; Vector generalized linear models; Skellam distribution; Ames test

LISTA DE FIGURAS

Figura 1 - Passos para execução do teste de <i>Salmonella</i> /microsossoma	28
Figura 2 - Foto de uma réplica do controle negativo	30
Figura 3 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_0$ do preditor de Bernstein	83
Figura 4 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_1$ do preditor de Bernstein	84
Figura 5 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_0$ do preditor de Breslow	85
Figura 6 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_1$ do preditor de Breslow	86
Figura 7 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\gamma}$ do preditor de Breslow	87
Figura 8 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_0$ do preditor de Krewski	88
Figura 9 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_1$ do preditor de Krewski	89
Figura 10 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\gamma}$ do preditor de Krewski	90
Figura 11 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_0$ do preditor de Margolin	91
Figura 12 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_1$ do preditor de Margolin	92
Figura 13 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\gamma}$ do preditor de Margolin	93
Figura 14 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_0$ do preditor de Myers	94
Figura 15 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_1$ do preditor de Myers	95
Figura 16 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\gamma}$ do preditor de Myers	96
Figura 17 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_0$ do preditor de Stead	97
Figura 18 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_1$ do preditor de Stead	98
Figura 19 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_2$ do preditor de Stead	99
Figura 20 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\gamma}$ do preditor de Stead	100
Figura 21 - Gráficos de dispersão dos dados observados nos experimentos com a linhagem TA98 no ensaio de <i>Salmonella</i> /microsossoma	109
Figura 22 - Gráficos de dispersão dos dados observados nos experimentos com a linhagem TA100 no ensaio de <i>Salmonella</i> /microsossoma	110
Figura 23 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Krewski e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsossoma (TA98) realizado com o inseticida Fipronil, na ausência da fração microsossomal S9	114

Figura 24 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Krewski e componente aleatório quase-Poisson, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Fipronil, na ausência da fração microsomal S9	115
Figura 25 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Margolin e componente aleatório quase-Poisson, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Fipronil, na ausência da fração microsomal S9	116
Figura 26 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Myers e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Fipronil, na ausência da fração microsomal S9	117
Figura 27 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Myers e componente aleatório quase-Poisson, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Fipronil, na ausência da fração microsomal S9	118
Figura 28 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Krewski e componente aleatório Poisson, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Fipronil, na presença da fração microsomal S9	120
Figura 29 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Margolin e componente aleatório Poisson, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Fipronil, na presença da fração microsomal S9	121
Figura 30 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Myers e componente aleatório Poisson, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Fipronil, na presença da fração microsomal S9	122
Figura 31 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Bernstein e componente aleatório Poisson, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Tiametoxam, na ausência da fração microsomal S9	124
Figura 32 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Stead e componente aleatório Poisson, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Tiametoxam, na ausência da fração microsomal S9	125
Figura 33 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Stead e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Tiametoxam, na presença da fração microsomal S9	127
Figura 34 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Stead e componente aleatório quase-Poisson, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Tiametoxam, na presença da fração microsomal S9	128
Figura 35 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Margolin e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Fipronil, na ausência da fração microsomal S9	130

Figura 36 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Myers e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Fipronil, na ausência da fração microsomal S9	131
Figura 37 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Margolin e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Fipronil, na presença da fração microsomal S9	133
Figura 38 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Margolin e componente aleatório quase-potência, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Fipronil, na presença da fração microsomal S9	134
Figura 39 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Myers e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Fipronil, na presença da fração microsomal S9	135
Figura 40 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Myers e componente aleatório quase-potência, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Fipronil, na presença da fração microsomal S9	136
Figura 41 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Breslow e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência da fração microsomal S9	138
Figura 42 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Stead e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência da fração microsomal S9	139
Figura 43 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Stead e componente aleatório quase-Poisson, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência da fração microsomal S9	140
Figura 44 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Breslow e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Tiametoxam, na presença da fração microsomal S9	142
Figura 45 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Breslow e componente aleatório quase-potência, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Tiametoxam, na presença da fração microsomal S9	143
Figura 46 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Margolin e componente aleatório quase-potência, pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Tiametoxam, na presença da fração microsomal S9	144
Figura 47 - Resíduos e diagnósticos para o MNLGV da família binomial negativa com preditor de Myers pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Fipronil, na ausência e na presença da fração microsomal S9	150

Figura 48 - Resíduos e diagnósticos para o MNLGV da família binomial negativa com preditor de Myers pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Fipronil, na ausência e na presença da fração microsomal S9	151
Figura 49 - Resíduos e diagnósticos para o MNLGV da família binomial negativa com preditor de Breslow pré-selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência e na presença da fração microsomal S9	152
Figura 50 - Resíduos e diagnósticos para o MLGV pré-selecionado da família Skellam para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Tiametoxam, na ausência e na presença da fração microsomal S9	153
Figura 51 - Resíduos e diagnósticos para o MLGV pré-selecionado da família Skellam para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência e na presença da fração microsomal S9	154
Figura 52 - Valores ajustados e bandas de confiança do modelo final selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Fipronil, na ausência e na presença da fração microsomal S9 .	158
Figura 53 - Valores ajustados e bandas de confiança dos modelo finais selecionados para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Tiametoxam, na ausência e na presença da fração microsomal S9	159
Figura 54 - Valores ajustados e bandas de confiança do modelo final selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Fipronil, na ausência e na presença da fração microsomal S9 .	159
Figura 55 - Valores ajustados e bandas de confiança do modelo final selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência e na presença da fração microsomal S9	160
Figura 56 - Flexibilidade dos preditores tradicionais	180

LISTA DE TABELAS

Tabela 1 - Revertentes espontâneos em diferentes linhagens de <i>S. typhimurium</i> . .	29
Tabela 2 - Modelos propostos para análise dos dados provenientes do Teste de Ames (modelos tradicionais)	34
Tabela 3 - Algumas funções de variância para dados de contagem, organizadas pela família e método de estimação	45
Tabela 4 - Força de evidência do AIC	52
Tabela 5 - Dados provenientes dos experimentos com a linhagem TA98	56
Tabela 6 - Dados provenientes dos experimentos com a linhagem TA100	57
Tabela 7 - Valores dos parâmetros das simulações	60
Tabela 8 - Frequência média do sucesso de convergência do algoritmo de estimação para cada preditor tradicional utilizando os métodos de obtenção de valores iniciais, por sistema de equações ou por ajustes de submodelos, na ausência e na presença de perturbações	82
Tabela 9 - Testes escore de Dean para superdispersão, de acordo com cada preditor tradicional, para os dados do experimento de <i>Salmonella</i> /microsoma com a linhagem TA98, utilizando o inseticida Fipronil, na presença e na ausência de um sistema de ativação metabólica	111
Tabela 10 - Testes escore de Dean para superdispersão, de acordo com cada preditor tradicional, para os dados do experimento de <i>Salmonella</i> /microsoma com a linhagem TA98, utilizando o inseticida Tiametoxam, na presença e na ausência de um sistema de ativação metabólica	111
Tabela 11 - Testes escore de Dean para superdispersão, de acordo com cada preditor tradicional, para os dados do experimento de <i>Salmonella</i> /microsoma com a linhagem TA100, utilizando o inseticida Fipronil, na presença e na ausência de um sistema de ativação metabólica	112
Tabela 12 - Testes escore de Dean para superdispersão, de acordo com cada preditor tradicional, para os dados do experimento de <i>Salmonella</i> /microsoma com a linhagem TA100, utilizando o inseticida Tiametoxam, na presença e na ausência de um sistema de ativação metabólica	112
Tabela 13 - Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de <i>Salmonella</i> /microsoma utilizando o inseticida Fipronil e S9(−), com a linhagem TA98	113
Tabela 14 - Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de <i>Salmonella</i> /microsoma utilizando o inseticida Fipronil e S9(+), com a linhagem TA98	119
Tabela 15 - Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de <i>Salmonella</i> /microsoma utilizando o inseticida Tiametoxam e S9(−), com a linhagem TA98	123
Tabela 16 - Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de <i>Salmonella</i> /microsoma utilizando o inseticida Tiametoxam e S9(+), com a linhagem TA98	126

Tabela 17 -Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de <i>Salmonella</i> /microsoma utilizando o inseticida Fipronil e S9(-), com a linhagem TA100	129
Tabela 18 -Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de <i>Salmonella</i> /microsoma utilizando o inseticida Fipronil e S9(+), com a linhagem TA100	132
Tabela 19 -Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de <i>Salmonella</i> /microsoma utilizando o inseticida Tiametoxam e S9(-), com a linhagem TA100	137
Tabela 20 -Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de <i>Salmonella</i> /microsoma utilizando o inseticida Tiametoxam e S9(+), com a linhagem TA100	141
Tabela 21 -Testes de razão de verossimilhanças para os MNGLVs encaixados da família binomial negativa tipo-2 com preditor de Myers ajustados aos dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Fipronil, na ausência e na presença da fração microsomal S9 . .	148
Tabela 22 -Testes de razão de verossimilhanças para os MNGLVs encaixados da família binomial negativa tipo-2 com preditor de Myers ajustados aos dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Fipronil, na ausência e na presença da fração microsomal S9 .	149
Tabela 23 -Testes de razão de verossimilhanças para os MNGLVs encaixados da família binomial negativa tipo-2 com preditor de Breslow ajustados aos dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência e na presença da fração microsomal S9	149
Tabela 24 -Testes de razão de verossimilhanças para os MLGVs encaixados da família Skellam ajustados aos dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Tiametoxam, na presença da fração microsomal S9	149
Tabela 25 -Testes de razão de verossimilhanças para os MLGVs encaixados da família Skellam ajustados aos dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência e na presença da fração microsomal S9	149
Tabela 26 -Estimativas, desvios-padrões e intervalos de confiança para os parâmetros do modelo final selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Fipronil, na ausência e na presença da fração microsomal S9	157
Tabela 27 -Estimativas, desvios-padrões e intervalos de confiança para os parâmetros do modelo final selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Tiametoxam, na ausência da fração microsomal S9	157
Tabela 28 -Estimativas, desvios-padrões e intervalos de confiança para os parâmetros do modelo final selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA98) realizado com o inseticida Tiametoxam, na presença da fração microsomal S9	158

Tabela 29 -Estimativas, desvios-padrões e intervalos de confiança para os parâmetros do modelo final selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Fipronil, na ausência e na presença da fração microsomal S9	158
Tabela 30 -Estimativas, desvios-padrões e intervalos de confiança para os parâmetros do modelo final selecionado para os dados do ensaio de <i>Salmonella</i> /microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência e na presença da fração microsomal S9	160
Tabela 31 -Estimativas de máxima verossimilhança das medidas de interesse biológico estimadas pelo modelo selecionado, de acordo com a linhagem de <i>S. typhimurium</i> , inseticida e sistema de ativação metabólica avaliados em cada experimento	161

1 INTRODUÇÃO

O teste de *Salmonella*/microsoma, conhecido também por Teste de Ames (TA), surgiu no contexto da avaliação de risco, desempenhando um importante papel na identificação de substâncias capazes de induzir mutações e danos no DNA. Esses químicos, conhecidos como agentes mutagênicos, podem provocar câncer ou causar problemas de fertilidade em diversos organismos, incluindo os humanos. Às vezes referido como “teste de mutagenicidade em *Salmonella*”, o TA foi especialmente projetado para detectar a mutagênese causada por químicos e, ao longo dos anos, conquistou uma posição de destaque na comunidade científica, nas organizações governamentais e na iniciativa privada, sendo utilizado para diversos fins, entre os quais podem ser citadas a liberação de medicamentos e de defensivos agrícolas e a avaliação da qualidade ambiental.

Os ensaios consistem, simplificadaamente, na exposição de linhagens específicas da bactéria *Salmonella typhimurium* a diferentes doses (ou concentrações) de um produto ou de uma amostra ambiental. Todas as linhagens são auxotróficas para histidina, isto é, apenas crescem e desenvolvem colônias na presença deste aminoácido. A versão mais utilizada do teste de Ames é a conhecida como “incorporação em placas”, na qual os experimentos são conduzidos em placas de Petri contendo meio de cultura deficiente em histidina. Dessa forma, só haverá colônias formadas, visíveis a olho nu, nas placas onde as unidades formadoras de colônia (UFCs) tiveram alteração em seu DNA e se tornaram capazes de se desenvolver mesmo na ausência de histidina. Chamam-se a essas colônias de “revertentes” e, assim, a variável resposta nesses casos é o número de revertentes por placa, sendo já bastante conhecido o número de revertentes espontâneos. Note que a simplicidade da variável resposta é apenas aparente: as UFCs respondem de maneira diferente a uma dada dose de uma substância em particular, de forma que os efeitos de mutagenicidade e toxicidade estão confundidos – enquanto o primeiro fenômeno aumenta o número de revertentes em uma placa, o segundo o diminui – e é tarefa do estatístico solucionar essa situação controversa.

Algumas substâncias são inativas, isto é, não-mutagênicas em sua fórmula padrão, porém, após entrar nos organismos, particularmente nos mamíferos, elas são metabolizadas e, então, seus subprodutos podem apresentar efeitos deletérios. Uma vez que as bactérias não produzem as mesmas enzimas que os mamíferos para a metabolização de substâncias, os experimentos são geralmente conduzidos na presença e na ausência do sistema de ativação metabólica S9 (SAM S9, sistema S9), que é produzido em órgãos como o fígado. Assim, é de notável interesse biológico e estatístico avaliar a existência da interação entre o SAM S9 e as concentrações dos químicos em questão.

Os termos dessa possível interação, entretanto, não estão explicitamente incluídos nos diversos modelos propostos na literatura para a análise de dados do TA (modelos tradicionais). Esses modelos estão todos baseados nas mesmas suposições probabilísticas e

estão inseridos numa classe mais ampla de modelos, os modelos não-lineares generalizados (ou modelos não-lineares da família exponencial). Eventualmente, essa abordagem não produz bons ajustes aos dados e a busca de novos modelos se faz necessária.

A regressão de Skellam, uma distribuição oriunda da diferença de variáveis aleatórias Poisson independentes, é promissora, pois é capaz de separar os efeitos de toxicidade e mutagenicidade. Um modelo desse tipo pode ser construído dentro do contexto dos modelos lineares generalizados vetoriais, que permitem elaborar vetores de preditores lineares, um para cada parâmetro de uma dada distribuição. Assim, uma forma mais genérica surge imediatamente: os modelos não-lineares generalizados vetoriais, capazes, então, de incluir tanto os modelos tradicionais quanto os modelos da família Skellam como casos particulares.

Outro caso particular de interesse é o modelo não-linear generalizado da família binomial negativa. Historicamente, esta distribuição tem sido usada para tratar, em dados de contagem, da superdispersão. Este fenômeno foi extensamente debatido na literatura estatística, e, apesar de linhas gerais terem sido muito bem estabelecidas, existem obstáculos para o desenvolvimento e aplicação de critérios objetivos para detectar a presença da superdispersão, determinar uma forma adequada da variância e, assim, selecionar o melhor modelo.

Os princípios de seleção de modelos estão baseados em testes de bondade de ajuste e critérios de informação, e, em ambos os casos, a aplicação de análise de resíduos e diagnósticos. Estas análises dependem do contexto estatístico, isto é, da classe de modelos que são ajustados aos dados. No intuito de criar uma plataforma capaz de integrar, não apenas resíduos e diagnósticos, mas também estimação e seleção, um dos objetivos deste trabalho foi desenvolver um pacote transparente, intuitivo e de fácil acesso para que cientistas, naturais ou de dados, possam analisar dados provenientes do ensaio de *Salmonella*/microsoma.

Este ensaio foi realizado no presente trabalho, utilizando, na presença e na ausência do SAM S9, diferentes concentrações de dois inseticidas, Tiametoxam e Fipronil, que são de interesse agrônomo e são amplamente empregados nas lavouras do Brasil. Assim, um segundo objetivo deste trabalho foi encontrar um modelo com bom ajuste aos dados provenientes do experimento desenvolvido.

O presente trabalho está organizado em seções, mas segue o molde tradicional. A revisão da literatura (seção 2) inclui elementos básicos de caráter biológico do teste de *Salmonella*/microsoma e introduz os modelos já propostos para a análise de dados (seções 2.1 e 2.2). Como esses modelos são casos particulares de um modelo não-linear generalizado, a seção 2 aborda sua formulação geral, a estimação, passando por análise de resíduos e questões sobre superdispersão (seções 2.3.1 a 2.3.4). A revisão prossegue com os modelos generalizados vetoriais na seção 2.4, já que os modelos da família Skellam são casos particulares e, como citado, se apresentam promissores. As questões sobre seleção

de modelos mais pertinentes a este trabalho são abordadas na seção 2.5, finalizando a seção 2. Posteriormente, tem-se o material (seção 3), que conta com uma breve descrição do experimento realizado neste trabalho e os dados obtidos. A seguir, na metodologia, além das particularidades dos algoritmos de estimação e dos estudos de simulação efetuados, merecem destaque a seção 4.3, que aborda o modelo não-linear generalizado vetorial, tratando-se de uma generalização imediata, mas inédita, dos modelos lineares generalizados vetoriais; e a 4.5, que trata mais detalhadamente das questões computacionais e do desenvolvimento do pacote. Já a seção 5 está dividida em resultados dos modelos tradicionais, resultados dos modelos propostos e seleção, incluindo as respectivas discussões sobre os temas pertinentes. Por fim, na conclusão deste trabalho, foram incluídos os modelos selecionados e suas implicações biológicas e as contribuições, perspectivas e pendências contextualizadas nas áreas da Estatística e da Ecotoxicologia.

2 REVISÃO BIBLIOGRÁFICA

2.1 Teste de *Salmonella*/microsoma

A mutagenicidade pode ser definida como uma alteração permanente no conteúdo do material genético de um organismo e pode ser encontrada no nível de genes individuais, de grupos de genes ou até de cromossomos inteiros. A identificação de substâncias capazes de induzir mutações é de grande importância para as seguranças humana e ambiental, já que tais mudanças genéticas podem provocar câncer, comprometer as estratégias reprodutivas de organismos ou a viabilidade de comunidades biológicas inteiras. Apesar das diferenças nos mecanismos de reparo de ácidos nucleicos (TAFURT; MARIN, 2014), há semelhanças interespecíficas nas estruturas do DNA e do RNA e em outros processos fisiológicos que influenciam na ação mutagênica (DEARFIELD et al., 2002).

Nesse contexto, a avaliação do potencial mutagênico de químicos torna-se uma alternativa bastante plausível utilizando, em conjunto com outros organismos-testes, bioensaios com microorganismos. O TA, um dos métodos empregados para detectar esse risco, é um ensaio de curto prazo amplamente aceito na comunidade científica e nas iniciativas pública e privada. O teste faz uso de algumas linhagens de *S. typhimurium* com mutações pré-existentes, que não permitem que as bactérias sintetizem um aminoácido essencial, a histidina, e, portanto, não têm a capacidade de crescer e formar colônias na ausência dele. Entretanto, novas mutações nesses genes onde ocorrem as mutações pré-existentes, ou em genes próximos, podem reestabelecer a capacidade de sintetizar histidina. Estes novos mutantes podem crescer na ausência do aminoácido e formar colônias. As mutações em algumas das linhagens de *S. typhimurium* utilizadas no teste de Ames são constituídas pela alteração de um par de base, enquanto que, em outras linhagens, as mutações pré-existentes são de natureza múltipla, podendo apresentar adição ou deleção de mais pares de base. Sabendo os tipos de mutações pré-existentes, pode-se, então, inferir a respeito dos mecanismos de ação mutagênica de uma substância (MORTELMANS; ZEIGER, 2000).

Os procedimentos para a montagem da versão de “incorporação em placas” do ensaio, primeiramente proposta em Ames et al. (1973), Ames, Lee e Durston (1973), estão apresentados na Figura 1. Estes podem ser resumidos como segue: misturar, se presente, 50 µL do sistema S9 com 50 µL do produto sob investigação e acrescentar à solução contendo as bactérias dependentes de histidina (50 a 100 µL); adicionar essa mistura ao ágar de superfície (2 mL), contendo biotina e traço de histidina (0,05 mM de cada); o volume final deve então ser agitado e colocado sobre o meio de cultura mínimo em placas de Petri. Depois da solidificação do ágar de superfície, as placas devem ser invertidas e incubadas por 48 h a 37 °C. Após esse intervalo de tempo, o número de revertentes é contado. Podem ser visualizadas na Figura 2, extraída de Mortelmans e Zeiger (2000), as colônias de revertentes formadas em uma réplica do controle negativo (dose zero) da linhagem TA100.

Figura 1 - Diagrama apresentando os passos para execução do teste de Ames *Salmonella* (versão de “incorporação em placas”). A: 50 a 100 μL de solução contendo bactérias *S. typhimurium* auxotróficas para histidina; B: 50 μL da substância de interesse ou apenas solvente; C: 50 μL do sistema de ativação metabólica; D: 2 mL de ágar de superfície, contendo traço de histidina (0,05 mM); E: placa contendo meio de cultura mínimo; F: incubação por 48 h a 37 °C; G: contagem do número de revertentes. E.1 e G.1: visões superiores dos estágios E e G. Note que o volume final de A+B+C não deve exceder 200 μL . Modificado de Mortelmans e Zeiger (2000)

Segundo Mortelmans e Zeiger (2000), a pequena quantidade de histidina presente no ágar de superfície é necessária para que todas as bactérias na placa possam sofrer algumas divisões celulares, um processo indispensável para que a mutagênese ocorra, de forma que os revertentes sejam facilmente identificados em contraste ao crescimento devido ao traço de histidina. O número de revertentes espontâneos é relativamente constante para cada linhagem da bactéria, variando um pouco de acordo com o nível do sistema S9 (ver Tabela 1).

Alguns químicos carcinogênicos, como aminas aromáticas e hidrocarbonetos aromáticos policíclicos, são biologicamente inativos, a não ser que sejam metabolizados para formas ativas. Nos humanos e em outros animais, o citocromo P450, um sistema de oxidação metabólica presente principalmente no fígado, nos pulmões e nos rins, é capaz de metabolizar uma grande parte desses compostos para formas reativas, interagentes com o DNA. Esses subcompostos, metabólitos intermediários, são mutagênicos em potencial e, uma vez que as bactérias não possuem essa capacidade metabólica, um sistema de ativação metabólica exógeno, tipicamente mamífero, deve ser adicionado ao ensaio. Esse sistema de ativação metabólica geralmente consiste de uma fração homogênea de fígado de rato, conhecida como fração microssomal S9. Mais detalhes da biologia e da bioquímica

envolvidas no TA são apresentados por Mortelmans e Zeiger (2000).

Tabela 1 - Número de revertentes espontâneos de acordo com a linhagem de *S. typhimurium*, considerando a presença (+) ou a ausência (−) do sistema de ativação metabólica S9. Adaptado de Mortelmans e Zeiger (2000)

Linhagem	Número de reverentes					
	S9(−)			S9(+)		
TA97	75	a	200	100	a	200
TA98	20	a	50	20	a	50
TA100	75	a	200	75	a	200
TA102	100	a	300	200	a	400
TA104	200	a	200	300	a	400
TA1535	5	a	20	5	a	20
TA1537	5	a	20	5	a	20
TA1538	5	a	20	5	a	20

O protocolo completo, de domínio público, para o desenvolvimento do TA teve sua versão nacional desenvolvida pela Sociedade Brasileira de Mutagênese, Carcinogênese e Teratogênese Ambiental (SBMCTA) e está disponível em Umbuzeiro et al. (2004). As recomendações de cunho estatístico presentes nesse protocolo, que parecem carecer de embasamento teórico adequado, incluem o planejamento do experimento, o ajuste de um modelo aos dados e os testes de hipóteses.

Quanto ao delineamento proposto no referido protocolo, é recomendada a utilização de, pelo menos, 5 doses (ou concentrações) da substância de interesse com três repetições por dose ou 7 doses (ou concentrações) com duas repetições por dose. Após a execução do experimento, é apresentada a seguinte série de passos para análise dos dados:

- (i) Aplicar uma transformação da família Box-Cox aos dados, com potência 1,5 e, então, proceder a uma análise de variância, considerando as doses como níveis qualitativos de um fator, separadamente para cada sistema S9 e para cada substância ou amostra ambiental;
- (ii) Ajustar um dos modelos apresentados na seção 2.2 para avaliar o efeito de dose sobre a resposta, também separadamente para cada sistema S9 e para cada substância ou amostra ambiental;
- (iii) Calcular o índice de mutagenicidade

$$IM_i = \frac{\bar{y}_i}{\bar{y}_c}, \quad (1)$$

em que $\bar{y}_i$ é o número médio de revertentes para a i -ésima dose ($i = 1, 2, \dots, k$) e $\bar{y}_c$ é o número médio de revertentes observados no controle negativo, isto é, para a dose 0.

A conclusão da análise é obtida pela combinação dos passos supracitados, de forma que uma substância será considerada mutagênica se os resultados das análises atenderem a todos os seguintes quesitos: a hipótese $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ for rejeitada no teste F da análise de variância a um fator; o parâmetro associado ao potencial mutagênico em um dos modelos da seção 2.2 for importante para a qualidade de ajuste do modelo; e o $IM \geq 2$ em todas as k doses (para as linhagens TA98 e TA100) (UMBUZEIRO et al., 2004). O primeiro passo não está incluído nos protocolos internacionais e, com relação ao terceiro passo, Hamada, Wada e Sakamoto (1994) demonstram que é irrelevante, dado que exista um modelo que apresente bom ajuste a um determinado conjunto de dados.

Figura 2 - Foto extraída de Mortelmans e Zeiger (2000), que apresenta as colônias de revertentes espontâneos da linhagem TA100 em uma réplica do controle negativo (dose zero)

2.2 Modelos propostos para o teste de *Salmonella*/microssoma

Na literatura, foram propostos diversos modelos, tanto paramétricos quanto não-paramétricos, para a análise dos dados do TA. Este estudo não tem como foco métodos não-paramétricos, mas alguns exemplos desse tipo de abordagem estão disponíveis em Wahrendorf, Mahon e Schumacher (1985), Simpson e Margolin (1986) e em Simpson e Margolin (1990).

Todos os modelos apresentados nesta seção, vistos mais detalhadamente, são paramétricos e foram baseados em Eckardt e Haynes (1977). Estes autores observaram o número de colônias mutantes de levedura formado após a exposição das unidades experimentais (placas de Petri, contendo meio de cultura mínimo e UFCs) a diferentes níveis de radiação ultravioleta (UV). Baseando-se nessas observações, propuseram um modelo mecanicista, descrito a seguir em termos de probabilidade e estatística.

Seja M uma variável aleatória latente com distribuição exponencial, descrevendo a sensibilidade das UFCs à mutação por UV. Suponha também que o parâmetro dessa distribuição seja u , uma constante positiva relacionada ao potencial mutagênico da UV. Assim, a probabilidade de ocorrer mutação em uma UFC qualquer, exposta a uma dose específica x_i , pode ser escrita como

$$P(M \leq x_i) = \int_0^{x_i} u e^{-ut} dt = 1 - e^{-ux_i} = F(x_i), \quad (2)$$

uma vez que ocorre mutação se, e apenas se, a sensibilidade M for menor ou igual à dose de exposição x_i .

Seja, agora, D outra variável aleatória latente com distribuição exponencial, descrevendo a tolerância das UFCs à morte por UV. Suponha também que o parâmetro dessa distribuição seja v , uma constante positiva relacionada ao potencial letal da UV. Assim, a probabilidade de não ocorrer morte de uma UFC qualquer exposta a uma dose específica x_i pode ser escrita como

$$P(D > x_i) = 1 - \int_0^{x_i} v e^{-vt} dt = e^{-vx_i} = \bar{G}(x_i), \quad (3)$$

visto que ocorre sobrevivência se, e apenas se, a tolerância D for maior que a dose de exposição x_i .

Supondo, por fim, que os processos de mutação e de morte sejam independentes e que a sobrevivência seja a mesma para mutantes e não-mutantes, então a probabilidade de se efetivamente detectar um mutante na i -ésima dose x_i é dada pela probabilidade conjunta de ocorrer a mutação e de que o mutante sobreviva; das equações (2) e (3), isto é obtido por:

$$p_i = F(x_i) \cdot \bar{G}(x_i) = (1 - e^{-ux_i})e^{-vx_i}. \quad (4)$$

Fazendo uso desse resultado, Eckardt e Haynes (1977) finalmente propuseram que o número de revertentes observados em uma placa na i -ésima dose é uma variável com distribuição de Poisson, Y_i , com média $\mu_i = m^* p_i$, em que m^* é o número inicial de UFCs na placa – geralmente, $m^* = 10^8$.

Todos os modelos baseados na equação (4), supondo diferentes funções para os termos das exponenciais e diferentes distribuições para o componente aleatório (CA), são apresentados a seguir e estão resumidos na Tabela 2 (p. 34). Estudos básicos sobre a forma de cada preditor podem ser encontradas no Apêndice A. Note que esses modelos não admitem explicitamente a interação entre os níveis do SAM S9 e as doses, uma questão já levantada por Weinstein e Lewinson (1978), que fizeram a análise dos dados do teste de Ames considerando um delineamento inteiramente casualizado, em esquema de tratamentos fatorial, e um modelo linear simples.

Modelo de Margolin. Após a publicação de Eckardt e Haynes (1977), os autores Margolin, Kaplan e Zeiger (1981) propuseram o seguinte preditor para a média da variável resposta do TA:

$$\eta_i = m(1 - e^{-(\beta_0^* + \beta_1^* x_i)})e^{-\gamma x_i}, \quad (5)$$

em que $m = g(m^*)$ é o número inicial de UFCs na escala da função de ligação $g(\cdot)$, β_0^* é o parâmetro associado ao número de revertentes espontâneos, β_1^* é o parâmetro de interesse associado à mutagenicidade e γ é a taxa na qual ocorre o decaimento exponencial devido à toxicidade, tal que β_0^* , β_1^* e $\gamma \geq 0$. Posteriormente, em Kim e Margolin (1999), foi disponibilizado um *software* para o ajuste de um modelo com esse preditor, além de uma revisão acerca da modelagem estatística dos dados do TA.

Modelo de Myers. O preditor proposto por Myers et al. (1981) para a média de revertentes é baseada na equação (5), em que $F(x_i) = 1 - e^{-(\beta_0^* + \beta_1^* x_i)}$ é aproximada pela série de Taylor de primeira ordem ao redor de $x_i = -\beta_0^*/\beta_1^*$, ou seja,

$$\begin{aligned} \eta_i &= m(1 - e^{-(\beta_0^* + \beta_1^* x_i)})e^{-\gamma x_i}, \\ &\approx m(\beta_0^* + \beta_1^* x_i)e^{-\gamma x_i}. \end{aligned}$$

Então, fazendo $m\beta_k^* = \beta_k$, $k = \{0, 1\}$, tem-se:

$$\eta_i = (\beta_0 + \beta_1 x_i)e^{-\gamma x_i}, \quad (6)$$

em que β_0 é o número de revertentes espontâneos, β_1 é o parâmetro de interesse associado à mutagenicidade e γ é a taxa na qual ocorre o decaimento exponencial devido à toxicidade, tal que β_0 , β_1 e $\gamma \geq 0$. Os autores não especificaram, entretanto, a distribuição da variável resposta, mas, devido à presença de superdispersão, definiram a função de variância como sendo $V(\mu) = a\mu^b$, em que $a, b > 0$. A estimação proposta pelos autores foi obtida por mínimos quadrados ponderados e os códigos para tal estão disponíveis em Myers et al. (1981).

Modelo de Stead. Os autores Stead et al. (1981) não consideraram superdispersão e supuseram simplesmente distribuição Poisson para as contagens obtidas nos dados do TA, mas consideraram um parâmetro a mais na equação (6), ou seja:

$$\eta_i = (\beta_0 + \beta_1 x_i^{\beta_2})e^{-\gamma x_i}, \quad (7)$$

em que β_0 é o número de revertentes espontâneos, β_1 e β_2 são os parâmetros de interesse associados à mutagenicidade e γ é a taxa na qual ocorre o decaimento exponencial devido

à toxicidade, tal que β_0 , β_1 e $\gamma \geq 0$, e $\beta_2 \geq 1$. Note que se $\beta_2 = 1$ na equação (7), então o modelo se reduz àquele da equação (6). A estimação proposta pelos autores foi realizada por meio de uma modificação do método de Gauss-Newton, para a qual um *software* foi posteriormente disponibilizado por Claxton et al. (1995).

Modelo de Bernstein. A proposta de Bernstein et al. (1982) para análise dos dados do TA é a mais simplista dentre as vistas até aqui. Os autores supõem que, para doses suficientemente baixas, a toxicidade não ocorre, de forma que um preditor adequado para a média da variável resposta na dose i seja dado por:

$$\eta_i = \beta_0 + \beta_1 x_i, \quad (8)$$

em que β_0 é o número de revertentes espontâneos e β_1 é o parâmetro de interesse associado à mutagenicidade, tal que β_0 e $\beta_1 \geq 0$. Para tratar da eventual superdispersão, os autores propuseram que a variável resposta tivesse uma distribuição de Poisson generalizada (PG), como em Consul e Jain (1973), e uma função de variância dada por $V(\mu) = a\mu^b$, em que $a, b > 0$. Apesar das facilidades de se trabalhar com esse modelo em qualquer plataforma estatística, há um software disponível em Claxton et al. (1995).

Modelo de Breslow. O preditor para média da variável resposta proposto por Breslow (1984), que permite linearização, é dado por:

$$\eta_i = e^{\beta_0} (x_i + \delta)^{\beta_1} e^{-\gamma x_i}, \quad (9)$$

em que β_0 é proporcional ao número de revertentes espontâneos, β_1 é o parâmetro de interesse associado à mutagenicidade, γ é a taxa na qual ocorre o decaimento exponencial devido à toxicidade e δ é um valor positivo arbitrário, usualmente escolhido como sendo o valor da menor dose não-nula, satisfazendo β_0 , β_1 , e $\gamma \geq 0$, e $\delta > 0$. O autor realizou a estimação dos parâmetros desse modelo via máxima quase-verossimilhança (McCULLAGH; NELDER, 1989), especificando apenas a função de variância $V(\mu) = \mu(1 + c\mu)$, e forneceu um código em GLIM (BAKER; NELDER, 1978) para que esse ajuste fosse realizado.

Modelo de Krewski. O modelo proposto por Krewski et al. (1993) para a média de revertentes é uma generalização da equação (6), como segue:

$$\eta_i = (\beta_0 + \beta_1 x_i) e^{-\gamma x_i^\alpha}, \quad (10)$$

em que β_0 é o número de revertentes espontâneos, β_1 é o parâmetro de interesse associado à mutagenicidade e γ é a taxa na qual ocorre o decaimento exponencial devido à toxicidade, tal que β_0 , β_1 e $\gamma \geq 0$, e $\alpha \geq 1$. Os autores não especificaram a distribuição de Y ,

mas, devido à possibilidade de haver superdispersão, utilizaram duas funções de variância $V_1(\mu) = a\mu^b$ e $V_2(\mu) = \mu(1 + c\mu)$, em que $a, b, c > 0$. Note que se $\alpha = 1$, então o modelo se reduz àquele da equação (6). Além disso, α não é um parâmetro, mas uma potência arbitrária — Krewski et al. (1993) fizeram a opção por $\alpha = 2$. A estimação proposta pelos autores foi obtida por máxima quase-verossimilhança (WEDDERBURN, 1974) e um *software* para o ajuste foi posteriormente disponibilizado por Leroux e Krewski (1993).

Todos as referências feitas neste trabalho ao conjunto de propostas da literatura, constantes nesta seção (2.2), para a análise dos dados do TA serão chamadas de “tradicionais”, “modelos tradicionais”, “preditores tradicionais” ou outro termo similar.

Tabela 2 - Modelos propostos para análise dos dados provenientes do Teste de Ames (modelos tradicionais). O preditor de cada modelo é obtido pelo produto da variável *offset*, da acumulada F , que descreve a ação mutagênica, e da sobrevivência $\bar{G}$, denotando a ação tóxica. $\hat{L}$ e $V(\mu)$ são, respectivamente, a função de verossimilhança ou quase-verossimilhança e a função de variância adotadas por cada autor: BN-2 se refere à distribuição binomial negativa tipo-2, QP se refere à quase-Poisson e P à Poisson

Modelo	<i>Offset</i>	F	$\bar{G}$	$\hat{L}$	$V(\mu)$
Margolin	m	$1 - e^{-(\beta_0 + \beta_1 x)}$	$e^{-\gamma x}$	BN-2	$\mu(1 + \mu\kappa^{-1})$
Myers	1	$\beta_0 + \beta_1 x$	$e^{-\gamma x}$	QP	$\phi\mu^\tau$
Krewski	1	$\beta_0 + \beta_1 x$	$e^{-\gamma x^\alpha}$	QP	$\phi\mu^\tau$
Breslow	1	$e^{\beta_0}(x + \delta)^{\beta_1}$	$e^{-\gamma x}$	QP	$\mu(1 + \mu\kappa^{-1})$
Bernstein	1	$\beta_0 + \beta_1 x$	1	QP	$\phi\mu^\tau$
Stead	1	$\beta_0 + \beta_1 x^{\beta_2}$	$e^{-\gamma x}$	P	μ

2.3 Modelos não-lineares generalizados

Todos os modelos propostos para analisar os dados do TA, resumidamente apresentados na Tabela 2, são casos particulares de um modelo não-linear generalizado (MNLG). Para tanto, é necessário esclarecer quais são as suposições, a fórmula geral e os processos de estimação e convergência nesses modelos, bem como abordar as semelhanças e diferenças dos MNLGs e seu caso linear, o modelo linear generalizado (MLG), que incluem aspectos de superdispersão e as análises de resíduos e diagnósticos.

2.3.1 Suposições e fórmula geral

Se $Y_1, Y_2, \dots, Y_n$ são variáveis aleatórias independentes e cada Y_i apresenta uma densidade na forma

$$p(y_i; \lambda_i, \phi_i) = \exp\{\phi_i[y_i\lambda_i - b(\lambda_i)] + c(y_i, \phi_i)\}, \quad (11)$$

então se diz que $p(y_i; \lambda_i, \phi_i)$ pertence à família exponencial em sua forma canônica. Na equação (11), $E(Y_i) = \mu_i = b'(\lambda_i)$, $\text{Var}(Y_i) = \phi_i^{-1}V_i(\mu_i)$ e $V_i(\mu_i) \equiv V_i = d\mu_i/d\lambda_i$ é chamada

de função de variância. Assume-se que $\phi_i > 0$ para $i = 1, 2, \dots, n$, sendo conhecido ou desconhecido (CORDEIRO; PAULA, 1989). Na literatura, o mais comum é utilizar a letra θ para designar o parâmetro canônico (ou natural) das distribuições pertencentes à família exponencial; no entanto, neste trabalho, houve a substituição pela letra λ , apenas para evitar ambiguidade de notação nas seções subsequentes.

Um modelo não-linear generalizado pode ser, seguindo Turner e Firth (2007), genericamente representado por:

$$g[E(\mathbf{Y}) \equiv \boldsymbol{\mu}] = \boldsymbol{\eta} = f(\mathbf{X}, \boldsymbol{\theta}), \quad (12)$$

em que $\mathbf{Y}$ é o vetor do componente aleatório, cuja distribuição ou pertence à família exponencial (equação (11)) ou à família exponencial quadrática (ver Vonesh (2012, p. 152)); $g(\cdot)$ é a função de ligação, conhecida e diferenciável; $\boldsymbol{\eta}$ é o vetor dos preditores; $f(\mathbf{X}, \boldsymbol{\theta})$ é o componente determinístico do modelo – uma função tipicamente não-linear, contínua e diferenciável, que relaciona a matriz $\mathbf{X}$ de variáveis explanatórias ao vetor de parâmetros desconhecidos $\boldsymbol{\theta}$. Note que $\boldsymbol{\theta}$ não guarda qualquer relação com os parâmetros naturais da família exponencial.

Os MNLGs podem ser entendidos como uma extensão dos MLGs ou, mais apropriadamente, como uma combinação das teorias do modelo não-linear (MNL) (GALLANT, 1975) e dos MLGs (McCULLAGH; NELDER, 1989). Estes autores consideraram a presença de parâmetros extras que entraram no preditor de forma não-linear, entretanto, em todos os casos, eles optaram por uma aproximação linear, utilizando expansões em séries de Taylor de primeira ordem em torno do termo não linear e ponderaram: “é geralmente imprudente tentar incluir mais do que alguns poucos parâmetros não-lineares por esse método” (McCULLAGH; NELDER, 1989, p. 380).

Em contraste com os MLGs, os MNLGs não impõem a restrição de linearidade do preditor, mas conservam a estrutura de dependência da variância em relação à média, bem como as simplificações matemáticas que emergem da utilização das distribuições que pertencem à família exponencial e das funções de ligação canônicas.

Nos MNLGs, a distribuição associada à variável resposta pode pertencer à família exponencial quadrática, abordagem especialmente utilizada para modelos de séries temporais ou de respostas multivariadas com estrutura de dependência. No entanto, uma vez que este não é o tema do presente trabalho, a discussão sobre quaisquer outros aspectos da família exponencial quadrática é remetida aos trabalhos de Prentice (1988), Prentice e Zhao (1991) e Vonesh (2012).

Alguns autores, como Krewski e Franklin (1991) e Vonesh (2012), optaram por omitir a função de ligação da equação (12), fazendo $\boldsymbol{\mu} = g^{-1}[f(\mathbf{X}, \boldsymbol{\theta})] = h(\mathbf{X}, \boldsymbol{\theta})$. Ainda que este seja um caminho viável para a definição formal de um MNLG, nestes casos há perda de generalidade: $g(\cdot)$ é fixada ser ou a identidade ou a ligação canônica. Usando

uma notação semelhante, mas sem perda de generalidade, Wei (1998) define formalmente os modelos não-lineares da família exponencial.

Os autores Davidian e Giltinan (1995) argumentam que, em comparação aos MLGs, não há penalidade em admitir que $f(\mathbf{X}, \boldsymbol{\theta})$ seja não-linear.

Além do TA, os MNLGs são utilizados em diversas áreas do conhecimento, dentre as quais estão os trabalhos de Welch, Adams e Corman (1963 apud McCULLAGH; NELDER, 1989) nas Ciências Agrícolas, de Darby e Ellis (1976) na área da Saúde e de Cautres, Heath e Firth (1998) nas Ciências Sociais.

2.3.2 Estimação

Máxima verossimilhança em MNLGs. Os estimadores de máxima verossimilhança dos parâmetros de um MNLG podem ser obtidos pela combinação dos métodos de estimação dos MLGs e dos MNLs; similarmente, a solução das equações de verossimilhança pode ser obtida iterativamente e coincide com a solução de mínimos quadrados. Para tal, o algoritmo de mínimos quadrados reponderados iterativamente (MQRI) (JØRGENSEN, 1983), exaustivamente discutido em Green (1984), pode ser aplicado.

Partindo dos resultados de Jørgensen (1983, p. 27), a equação utilizando escore de Fisher para obter as estimativas de máxima verossimilhança dos parâmetros de um modelo regular é dada por:

$$\boldsymbol{\theta}^{(a+1)} = \{\tilde{\mathbf{X}}^{\top(a)} \mathbf{W}^{(a)} \tilde{\mathbf{X}}^{(a)}\}^{-1} \tilde{\mathbf{X}}^{\top(a)} \mathbf{W}^{(a)} \mathbf{z}^{(a)}, \quad (13)$$

em que $\boldsymbol{\theta}^{(a)}$ é o vetor de estimativas dos parâmetros do modelo na a -ésima iteração, $\tilde{\mathbf{X}} = \partial \boldsymbol{\eta} / \partial \boldsymbol{\theta}^{\top}$ é a matriz de derivadas primeiras do vetor de preditores em relação aos parâmetros do modelo, $\mathbf{W} = -E \left(\frac{\partial^2 L}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^{\top}} \right)$ é a matriz de informação esperada para os preditores e $\mathbf{z} = \tilde{\mathbf{X}} \boldsymbol{\theta} + \mathbf{W}^{-1} \frac{\partial L}{\partial \boldsymbol{\eta}}$ é a variável resposta modificada, cujo logaritmo da função de verossimilhança é denotado por $L \equiv L[\boldsymbol{\eta}(\boldsymbol{\theta}) \mid \mathbf{y}]$.

Para os MNLGs, a equação (13) pode ser simplificada, fazendo uso das suposições de que há independência entre as observações, de que a distribuição da variável resposta pertence à família exponencial e de que não há censura.

Seja $\mathbf{U}(\boldsymbol{\eta}) \equiv \mathbf{U}$ o vetor escore, de comprimento n , para os preditores $\boldsymbol{\eta}$ de uma amostra de n observações. Os elementos típicos de $\mathbf{U} = \partial L / \partial \boldsymbol{\eta}$, pela regra da cadeia, são dados por:

$$U_i = \frac{\partial L_i}{\partial \eta_i} = \frac{dL_i}{d\lambda_i} \frac{d\lambda_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} = \phi_i [y_i - b'(\lambda_i)] \frac{1}{V_i} \frac{d\mu_i}{d\eta_i}.$$

Assim, a matriz de informação esperada para $\boldsymbol{\eta}$ pode ser escrita como

$$\mathbf{W} = -E \left(\frac{\partial^2 L}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^{\top}} \right) = E \left(\frac{\partial L}{\partial \boldsymbol{\eta}} \frac{\partial L}{\partial \boldsymbol{\eta}^{\top}} \right) = E [\mathbf{U}(\boldsymbol{\eta}) \mathbf{U}(\boldsymbol{\eta})^{\top}]. \quad (14)$$

Note que a matriz (14) é diagonal $n \times n$, sob suposição de independência, cujos elementos típicos são da forma

$$\begin{aligned}
 E(U_i^2) &= E \left[\phi_i^2 (y_i - \mu_i)^2 \frac{1}{V_i^2} \left(\frac{d\mu_i}{d\eta_i} \right)^2 \right] \\
 &= \frac{\phi_i^2}{V_i^2} \left(\frac{d\mu_i}{d\eta_i} \right)^2 E[(y_i - \mu_i)^2] \\
 &= \frac{\phi_i^2}{V_i^2} \left(\frac{d\mu_i}{d\eta_i} \right)^2 \text{Var}(y_i) \\
 &= \frac{\phi_i}{V_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2.
 \end{aligned} \tag{15}$$

Sejam então $\omega_i = \frac{1}{V_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2$ os elementos de $\mathbf{\Omega} = \text{diag}\{\omega_1, \omega_2, \dots, \omega_n\}$ e, similarmente, $\mathbf{\Phi} = \text{diag}\{\phi_1, \phi_2, \dots, \phi_n\}$ – está claro que, de (14) e (15), $\mathbf{W} = \mathbf{\Omega}\mathbf{\Phi}$.

Da equação (13) pode-se ainda reescrever

$$\begin{aligned}
 \mathbf{z} &= \tilde{\mathbf{X}}\boldsymbol{\theta} + \mathbf{W}^{-1}\mathbf{U}(\boldsymbol{\eta}) \\
 &= \tilde{\mathbf{X}}\boldsymbol{\theta} + \mathbf{A}(\mathbf{y} - \boldsymbol{\mu}),
 \end{aligned}$$

em que $\mathbf{A} = \text{diag}\left\{\frac{d\eta_1}{d\mu_1}, \frac{d\eta_2}{d\mu_2}, \dots, \frac{d\eta_n}{d\mu_n}\right\}$. Por fim, o algoritmo simplificado, ideal para o ajuste de um MNLG aos dados do TA, é dado por:

$$\boldsymbol{\theta}^{(a+1)} = \left\{ \tilde{\mathbf{X}}^{\top(a)} \mathbf{\Omega}^{(a)} \mathbf{\Phi} \tilde{\mathbf{X}}^{(a)} \right\}^{-1} \tilde{\mathbf{X}}^{\top(a)} \mathbf{\Omega}^{(a)} \mathbf{\Phi} \mathbf{z}^{(a)}, \tag{16}$$

também assim descrito em Cordeiro e Paula (1989). Note que a suposição de independência ou, neste caso, equivalentemente de amostragem aleatória, é necessária para se obter estimativas válidas pelo algoritmo iterativo apresentado na equação (16). Para situações que violam esse pressuposto, os estimadores de máxima verossimilhança e o algoritmo adequado para a obtenção das estimativas devem ser realizadas como descrito por Wei (1998, p. 19–27).

Convergência. Todos os algoritmos iterativos de estimação requerem um critério de convergência, isto é, uma condição pré-estabelecida que, se verdadeira, determina a parada do algoritmo. Para os MNLGs, dois desses critérios, amplamente utilizados em MLGs, podem ser aplicados: (i) o algoritmo para se a soma dos módulos das diferenças entre os pares das estimativas dos parâmetros em duas iterações subseqüentes for menor que um valor suficientemente pequeno, ε_1 ; ou (ii) o algoritmo para se o módulo da diferença entre as funções *deviance* de duas iterações subseqüentes for menor que um valor suficientemente pequeno, ε_2 . Em termos matemáticos, (i) e (ii) podem ser respectivamente representados por:

- (i) $\|\boldsymbol{\theta}^{(a+1)} - \boldsymbol{\theta}^{(a)}\| < \varepsilon_1$;
- (ii) $|D(\mathbf{y}, \hat{\boldsymbol{\mu}}^{(a+1)}) - D(\mathbf{y}, \hat{\boldsymbol{\mu}}^{(a)})| < \varepsilon_2$.

A função *deviance* está definida na seção 2.3.3, pois também desempenha um importante papel na análise de resíduos.

2.3.3 Resíduos, diagnósticos e gráficos

Todos os tipos de resíduos apresentados nesta seção são baseados em Cordeiro e Demétrio¹ (em fase de elaboração), que, sem perda de generalidade, se aplicam tanto aos MLGs quanto aos MNLGs, como pode ser verificado em Wei (1998, p. 121, Tabela 6.3).

Medidas de *leverage*. Essas quantidades são obtidas pela matriz de projeção $\mathbf{H}$:

$$\mathbf{H} = \mathbf{W}^{1/2} \tilde{\mathbf{X}} \left\{ \tilde{\mathbf{X}}^\top \mathbf{W} \tilde{\mathbf{X}} \right\}^{-1} \tilde{\mathbf{X}}^\top \mathbf{W}^{1/2},$$

em que $\tilde{\mathbf{X}}$ e $\mathbf{W} = \boldsymbol{\Omega}\boldsymbol{\Phi}$ são calculadas na convergência do algoritmo. As medidas de *leverage* são os elementos da diagonal da matriz $\mathbf{H}$, satisfazendo $\text{tr}\{\mathbf{H}\} = p = \sum_{i=1}^n h_{ii}$. Isto implica que o valor médio de h_{ii} é p/n e, como cada h_{ii} representa o peso que i -ésima observação exerce sobre a estimativa $\hat{\mu}_i$, essas medidas são úteis para identificar pontos de alavanca e *outliers*. De modo geral, um gráfico de h_{ii} *versus* i (gráfico de índice) pode ser utilizado para determinar quais observações precisam de maior atenção, particularmente aquelas que apresentam valores de $h_{ii} > 2p/n$ (ver os trabalhos de Moolgavkar, Lustbader e Venzon (1984) e Wei (1998, p. 117) para algumas questões conflitantes sobre a definição da matriz de projeção em MNLGs).

Resíduos do tipo *working*.

$$r_i^W = (y_i - \hat{\mu}_i) \frac{d\hat{\eta}_i}{d\hat{\mu}_i}.$$

Os resíduos do tipo *working* servem de base para a obtenção de outros tipos de resíduos, mas, a princípio, não são de interesse direto.

Resíduos de Pearson.

$$r_i^p = \frac{y_i - \hat{\mu}_i}{\sqrt{V_i(\hat{\mu}_i)}},$$

tal que $X^2 = \sum_{i=1}^n (r_i^p)^2$ é a estatística qui-quadrado de Pearson.

Gráficos de resíduos de Pearson *versus* valores ajustados ($\hat{\mu}_i$), *versus* preditores ajustados ($\hat{\eta}_i$) e *versus* variáveis explanatórias (particularmente “doses”, para os dados do TA) são úteis para detectar potencial ausência de termos quadráticos ou de ordem

¹CORDEIRO, G.M.; DEMÉTRIO, C.G.B. **Modelos lineares generalizados e extensões.**

superior no preditor, problemas com a especificação da função de variância ou falta de independência entre as observações.

Para verificar autocorrelação, também podem ser construídos gráficos de defasagem, isto é r_i^p versus r_{i-c}^p , em que c é o tamanho da defasagem. Para ensaios do TA conduzidos com amostras ambientais ou tempos de biodegradação, a análise desse tipo de gráfico é muito importante.

Além desses gráficos, resíduos ordenados *versus* quantis da distribuição normal (*normal Q-Q plot*) podem ser úteis para identificar observações discrepantes. Também é comum o uso dos valores absolutos dos resíduos *versus* as estatísticas de ordem da distribuição normal (*half-normal plot*). Como, em geral, não há garantia de que a distribuição dos resíduos seja, de fato, normal, é recomendado o uso de envelopes simulados, que podem ser obtidos para quaisquer outras estatísticas de interesse, como medidas de *leverage*. Especificamente para MNLGs, o trabalho de Simas e Cordeiro (2009) traz alguns desenvolvimentos importantes com relação à distribuição dos resíduos de Pearson.

Resíduos de Pearson estudentizados.

$$r_i^{p*} = \frac{r_i^p}{\sqrt{1 - h_{ii}}},$$

em que h_{ii} é o i -ésimo elemento da diagonal da matriz de projeção $\mathbf{H}$. Os mesmos tipos de gráficos construídos para os resíduos de Pearson podem também ser construídos para os resíduos de Pearson estudentizados.

Componentes de *deviance*.

$$r_i^d = \text{sin}(\hat{y}_i - \hat{\mu}_i) \sqrt{2 [L(y_i; y_i) - L(\hat{\mu}_i; y_i)]}, \quad (17)$$

em que r_i^d é chamado componente de *deviance*. A soma dos quadrados dos componentes de *deviance* é chamada “função *deviance*”, dada por $D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n (r_i^d)^2$, que é uma medida de discrepância do ajuste do modelo aos dados observados, e é também usada como critério de convergência do algoritmo de estimação. Em alguns casos, a *deviance* escalonada, $\tilde{D} = \phi D$, é uma medida de interesse, particularmente naqueles em que o parâmetro de dispersão ϕ é desconhecido ou não fixado *a priori*.

Os gráficos para os componentes de *deviance* também podem ser construídos da mesma forma que para os resíduos de Pearson.

Componentes de *deviance* estudentizados.

$$r_i^{d*} = \frac{r_i^d}{\sqrt{1 - h_{ii}}} \quad (18)$$

em que h_{ii} é o i -ésimo elemento da diagonal da matriz de projeção $\mathbf{H}$. Também se aplicam os mesmos tipos de gráficos.

Medidas de influência e curvatura locais. As medidas de influências servem para identificar o efeito de perturbações na verossimilhança. Supondo um esquema de perturbação caso-a-caso, a verossimilhança fica:

$$L[\boldsymbol{\eta}(\boldsymbol{\theta}); \mathbf{v}] = \sum_{i=1}^n v_i L[\eta_i(\theta_i)],$$

tal que $v_i \in [0, 1]$. Seja $\mathbf{v}_0$ um vetor unidade de comprimento n que representa a ausência de perturbação, assim $L[\boldsymbol{\eta}(\boldsymbol{\theta})|\mathbf{v}_0] = L[\boldsymbol{\eta}(\boldsymbol{\theta})]$ (SVETLIZA; PAULA, 2003). Para simplicidade de notação, daqui para frente, a verossimilhança perturbada será denotada por $L_{\mathbf{v}}$ e a verossimilhança sem perturbação apenas por L . A ideia de influência local, apresentada por Cook (1986), consiste em caracterizar o comportamento de alguma medida de deslocamento em torno de $\mathbf{v}_0$.

Seja $\text{MD}(L, L_{\mathbf{v}})$ uma função que represente uma medida de deslocamento qualquer entre L e $L_{\mathbf{v}}$, que satisfaça $\text{MD}(L, L_{\mathbf{v}_0}) = 0$. Pode-se então calcular $\text{MD}(L, L_{\mathbf{v}_0 + a\mathbf{l}})$ para um conjunto de valores $a \in \mathbb{R}$ e uma direção $\mathbf{l}$ de norma unitária, de forma que, sob a suposição feita sobre $\text{MD}(\cdot)$, existe um mínimo local em $a = 0$.

Assim, sujeito à escolha de $\mathbf{l}$, podem ser obtidas, em torno $a = 0$, diferentes curvaturas normais. Como $\|\mathbf{l}\| = 1$, a curvatura normal é dada por $C_{\mathbf{l}} = 2|\mathbf{l}^\top \boldsymbol{\Delta}^\top \mathbf{K}^{-1} \boldsymbol{\Delta} \mathbf{l}|$, em que cada elemento de Δ_{ij} é dado pela derivada segunda da verossimilhança perturbada com relação ao parâmetro do preditor θ_i e à perturbação v_j , avaliados em $\hat{\theta}_i$ e v_{0j} ; e $\mathbf{K}$ é a matriz de informação observada na convergência. Sejam, então, $C_{\mathbf{l}_{\max}}$ e $\mathbf{l}_{\max}$ o maior auto-valor absoluto e o auto-vetor correspondente da matriz $\mathbf{B} = \boldsymbol{\Delta}^\top \mathbf{K}^{-1} \boldsymbol{\Delta}$, denotando, respectivamente, a maior curvatura e a direção em que essa curvatura ocorre. O gráfico de índices de $\mathbf{l}_{\max}$ pode revelar observações que são localmente influentes para o conjunto de estimativas $\hat{\boldsymbol{\theta}}$. Para investigar sub-conjuntos de $\hat{\boldsymbol{\theta}}$, pode-se seguir uma estratégia similar, definindo uma nova matriz $\mathbf{B}^* = \boldsymbol{\Delta}^\top (\mathbf{K}^{-1} - \mathbf{B}_1) \boldsymbol{\Delta}$, em que os elementos de $\mathbf{B}_1$ são iguais a zero para os parâmetros de interesse e recebem os elementos de $\mathbf{K}^{-1}$ nas posições dos demais parâmetros (COOK, 1986; SVETLIZA; PAULA, 2003; WEI, 1998, p. 124-128).

Outra abordagem foi proposta por Lesaffre e Verbeke (1998), que sugeriram adotar n direções diferentes, $\mathbf{l}_i$, e então verificar a curvatura $C_{\mathbf{l}_i}$ em cada uma delas. Os vetores $\mathbf{l}_i$ são compostos por zeros em todas as posições, exceto na i -ésima posição, que recebe 1. A norma de $\mathbf{l}$, assim, permanece igual a 1 e as curvaturas normais podem ser obtidas. Assim, um gráfico de índices de $C_{\mathbf{l}_i}$ revela curvaturas locais excessivas na direção de cada observação.

2.3.4 Superdispersão

Os dados provenientes do teste de *Salmonella*/microsoma, independente da forma sistemática escolhida, são constituídos por contagens e, dessa forma, o componente aleatório mais natural é Poisson. A distribuição Poisson impõe, entretanto, a igualdade entre média e variância. Essa restrição dificilmente tem embasamento empírico, de forma que, na maioria dos casos, a variância é maior que a média e, a esse fenômeno, chamamos “superdispersão”. O contrário, a subdispersão, também pode ocorrer, mas é muito menos frequente e, até o momento, nenhum caso de subdispersão foi encontrado nos dados do TA.

A falha em detectar e incorporar a sub ou superdispersão nos modelos tem consequências, que incluem a subestimação dos erros-padrões, inferência incorreta a respeito dos parâmetros do preditor e falta de ajuste. A busca por um melhor ajuste foi o que levou à escolha de diferentes componentes aleatórios nos modelos tradicionais, conforme explicitado na Tabela 2.

Há muitas propostas na literatura para tratar da superdispersão – desde modelos de mistura a modelos dedicados à dispersão (HOUGAARD; LEE; WHITMORE, 1997; JØRGENSEN, 1997). Entre os modelos de mistura, merecem atenção o de Poisson-Gaussiana Inversa (DEAN; LAWLESS; WILLMOT, 1989) e o de Poisson-Tweedie (KOKONENDJI; DOSSOU-GBÉTÉ; DEMÉTRIO, 2004), e, principalmente, a mistura Poisson-gama, que leva a um componente aleatório binomial negativo em MNLGs. Essa distribuição é obtida pela composição das distribuições Poisson e gama, como segue: sejam $Y_1, Y_2, \dots, Y_n$ variáveis aleatórias independentes tais que $Y_i \sim \text{Poisson}(\Lambda_i)$, então, se todo $\Lambda_i \sim \Gamma(k, \mu_i/k)$ para qualquer $i = 1, 2, \dots, n$, a distribuição marginal de Y_i é obtida por

$$\begin{aligned}
 p_{Y_i, \Lambda_i}(y_i; \mu_i, k) &= \int_0^{+\infty} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \frac{\lambda_i^{k-1} e^{-\frac{k\lambda_i}{\mu_i}}}{\left(\frac{\mu_i}{k}\right)^k \Gamma(k)} d\lambda_i \\
 &= \frac{k^k}{\mu_i^k y_i! \Gamma(k)} \int_0^{+\infty} \lambda_i^{y_i+k-1} \exp \left\{ -\lambda_i \left(1 + \frac{k}{\mu_i} \right) \right\} d\lambda_i \\
 &= \frac{\Gamma(y_i + k)}{y_i! \Gamma(k)} \left(\frac{k}{k + \mu_i} \right)^k \left(\frac{\mu_i}{k + \mu_i} \right)^{y_i}.
 \end{aligned} \tag{19}$$

A distribuição da equação (19) tem esperança e variância dadas, respectivamente, por μ_i e $\mu_i(1 + \mu_i/k)$, e é chamada binomial negativa tipo 2 (BN-2) com parâmetros μ_i e k . Para detalhes sobre a estimação do parâmetro k da BN-2, ver os trabalhos de Hinde e Demétrio (1998), Lawless (1987) e Nakashima (1997).

Também existe uma forma mais genérica da distribuição binomial negativa, BN-G,

que é escrita como:

$$p(y_i; \mu_i, k) = \frac{\Gamma(y_i + k\mu_i^{1-g})}{y_i! \Gamma(k\mu_i^{1-g})} \left(\frac{k + \mu_i^g}{k} \right)^{-k\mu_i^{1-g}} \left(\frac{k + \mu_i^g}{k} \right)^{-y_i},$$

em que g é um parâmetro funcional, que estabelece o grau do polinômio da função de variância. Assim, se $g = 0$, obtém-se a uma função de variância constante com $V_i = \mu_i(1 + 1/k) \equiv \phi\mu_i$ e a distribuição associada é a binomial negativa tipo 1 (BN-1); se $g = 1$ tem-se uma função de variância quadrática associada à BN-2; se $g = 2$ tem-se uma função cúbica e assim sucessivamente.

É notável a função de variância geral para dados de contagem apresentada por Hinde e Demétrio (1998), que fizeram:

$$\text{Var}(Y_i) = \mu_i(1 + \phi\mu_i^\tau),$$

cujos parâmetros de superdispersão, ϕ e τ , podem ser estimados por perfil de verossimilhança ou podem recuperar as distribuições BN-1 com $\tau = 0$ e BN-2 com $\tau = 1$ e, para ambas, $\phi = k^{-1}$. De fato, existem os modelos de regressão Hinde-Demétrio (KONENDJI; DEMÉTRIO; ZOCCHI, 2007) e também modelos baseados na distribuição Poisson generalizada (CASTILLO; PÉREZ-CASANY, 2005; CONSUL; JAIN, 1973; JOE; ZHU, 2005), ambos para dados de contagem com superdispersão, mas que, a princípio, não podem ser entendidos como MNLGs.

Para MNLGs discretos, o parâmetro de dispersão $\phi = 1$ é fixado *a priori*. Assim, tanto para permitir ϕ desconhecido quanto para permitir parâmetros desconhecidos dentro das funções de variância, são necessários métodos de estimação modificados. Dentre esses métodos, podem ser citados a máxima quase-verossimilhança (MQV) (WEDDERBURN, 1974), a máxima quase-verossimilhança estendida (MQVE) (NELDER; PREGIBON, 1987) e a máxima pseudo-verossimilhança (MPV) (BRESLOW, 1990; DAVIDIAN; CARROLL, 1987). Estas três abordagens são comparadas no trabalho de Nelder e Lee (1992).

Método da máxima quase-verossimilhança. Seja l o logaritmo da função de quase-verossimilhança para $\boldsymbol{\mu}$, baseado nas n observações $\mathbf{y}$, tal que:

$$l(\boldsymbol{\mu}; \mathbf{y}) = \sum_{i=1}^n q_i(\boldsymbol{\mu}; \mathbf{y}),$$

em que

$$q_i(\boldsymbol{\mu}; \mathbf{y}) = \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi_i^{-1} V_i(t)} dt,$$

sob suposição de que os dois primeiros momentos populacionais das variáveis aleatórias $\mathbf{Y}$ que originaram as observações $\mathbf{y}$ são dados respectivamente por $E(\mathbf{Y}) = \boldsymbol{\mu}$ e $\text{Var}(\mathbf{Y}) =$

$\Phi \mathbf{V}(\boldsymbol{\mu})$, sendo $\mathbf{V}(\boldsymbol{\mu}) \equiv \mathbf{V} = \text{diag} \{V_i\}$.

Assim, o vetor quase-escore dos preditores pode ser escrito como

$$\frac{\partial l}{\partial \boldsymbol{\eta}} = \frac{\partial l}{\partial \boldsymbol{\mu}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}. \quad (20)$$

É fácil ver que $\partial l / \partial \boldsymbol{\mu} = \Phi \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu})$, já que, sob independência, um elemento típico é escrito como

$$\frac{\partial l}{\partial \mu_i} = \frac{\partial}{\partial \mu_i} \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi_i^{-1} V_i(t)} dt = \frac{\phi_i}{V_i} (y_i - \mu_i).$$

Como $\partial \boldsymbol{\mu} / \partial \boldsymbol{\eta}$ é diagonal $n \times n$ com elementos típicos $d\mu_i / d\eta_i$, o vetor quase-escore na equação (20) fica

$$\frac{\partial l}{\partial \boldsymbol{\eta}} = \mathbf{A}^{-1} \Phi \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}), \quad (21)$$

de forma que pode-se obter o vetor quase-escore para os parâmetros do preditor apenas pré-multiplicando, na equação (21) a matriz transposta de primeiras derivadas parciais dos preditores $\boldsymbol{\eta}$ em relação a $\boldsymbol{\theta}$ de dimensões $p \times n$:

$$\frac{\partial l}{\partial \boldsymbol{\theta}} = \left(\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}^\top} \right)^\top \left(\frac{\partial l}{\partial \boldsymbol{\eta}} \right) = \left(\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}^\top} \right)^\top \mathbf{A}^{-1} \Phi \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = \tilde{\mathbf{X}}^\top \mathbf{A}^{-1} \Phi \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}). \quad (22)$$

Rearranjando a equação (22) para estabelecer uma clara relação com o método da máxima verossimilhança, pode-se obter:

$$\frac{\partial l}{\partial \boldsymbol{\theta}} = \mathbf{U}(\boldsymbol{\theta}) = \tilde{\mathbf{X}}^\top \Phi \Omega \mathbf{A}(\mathbf{y} - \boldsymbol{\mu}),$$

pois $\Omega \mathbf{A} \equiv \mathbf{A}^{-1} \mathbf{V}^{-1}$. Assim como no método da máxima verossimilhança, a matriz de informação esperada também é o negativo do valor esperado de $\partial \mathbf{U}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$, dada por

$$\mathbf{i}(\boldsymbol{\theta}) = \tilde{\mathbf{X}}^\top \Phi \Omega \tilde{\mathbf{X}}, \quad (23)$$

e se Φ for um escalar fixo desconhecido, a matriz de covariâncias dos parâmetros do modelo pode ser escrita como $\text{Cov}(\boldsymbol{\theta}) = \phi^{-1} \left\{ \tilde{\mathbf{X}}^\top \Omega \tilde{\mathbf{X}} \right\}^{-1}$.

Tendo o vetor escore (equação (15)) e a matriz de informação esperada dos parâmetros do modelo (equação (23)), um processo iterativo para obter as estimativas $\hat{\boldsymbol{\theta}}$ é dado por:

$$\hat{\boldsymbol{\theta}}^{(a+1)} - \hat{\boldsymbol{\theta}}^{(a)} = \left\{ \left[\tilde{\mathbf{X}}^{\top(a)} \Phi \Omega^{(a)} \tilde{\mathbf{X}}^{(a)} \right]^{-1} \tilde{\mathbf{X}}^{\top(a)} \Phi \Omega^{(a)} \mathbf{A}^{(a)} \right\} (\mathbf{y} - \boldsymbol{\mu}),$$

que, após alguma manipulação algébrica, é um algoritmo equivalente àquele da equação

(16).

Um critério de convergência para esse algoritmo é baseado na função de quase-*deviance*, cujos componentes podem ser utilizados na análise de resíduos do modelo e são obtidos por:

$$d_i^q = -2 \int_{y_i}^{\mu_i} \frac{y_i - t}{V(t)} dt, \quad (24)$$

independente de Φ . Por outro lado, se $\Phi = \phi \mathbf{I}$, em que $\mathbf{I}$ é a matriz identidade, e ϕ precisa ser estimado, é possível utilizar o método dos momentos, baseado na estatística generalizada de Pearson, pois, para esse parâmetro, os comportamentos dos logaritmos das funções de verossimilhança e quase-verossimilhança são diferentes (McCULLAGH; NELDER, 1989, p. 328), como segue:

$$\hat{\phi}^{-1} = \frac{1}{n-p} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V_i(\hat{\mu}_i)}. \quad (25)$$

O método da equação (25) pode ser usado apenas se $V_i(\hat{\mu}_i)$, para $i = 1, \dots, n$, estiver livre de outras quantidades desconhecidas. Alguns casos em que outros parâmetros estão presentes em $\mathbf{V}$ e requerem estimação, o método da máxima quase-verossimilhança estendida (NELDER; PREGIBON, 1987; NELDER; LEE, 1992) pode ser aplicado.

Método da máxima quase-verossimilhança estendida. Está claro que as mesmas ideias para ajustar modelos por máxima verossimilhança (MV) são usadas para ajustar modelos por máxima quase-verossimilhança (MQV), fazendo apenas suposições sobre os dois primeiros momentos da distribuição dos dados. Entretanto, como a função *deviance* e a estatística de Pearson são medidas obtidas via a função de variância, não é possível avaliar a qualidade de ajuste para diferentes funções de variâncias, mantendo os preditores fixos.

Então, para possibilitar a comparação de diferentes funções de variância e permitir o uso de parâmetros extras nessas funções, Nelder e Pregibon (1987) propuseram a seguinte quantidade a ser maximizada a cada iteração do algoritmo (16):

$$l^+(\boldsymbol{\mu}; \mathbf{y}) = \sum_{i=1}^n q_i^+(\mu_i; y_i); \quad q_i^+(\mu_i; y_i) = -\frac{1}{2} \log [2\pi\phi^{-1}V(y_i)] - \frac{\phi}{2} d_i^q, \quad (26)$$

em que d_i^q é o mesmo da equação (24). Cordeiro e Demétrio (2008) discutiram propriedades desses estimadores e propuseram correções de vies.

A máxima quase-verossimilhança estendida (MQVE) permite estimar parâmetros desconhecidos tanto na função de ligação, quanto na função de variância por meio de perfis de verossimilhança, ou, mais apropriadamente neste caso, perfis de quase-verossimilhança estendida. Os métodos para obtenção dos perfis são, do ponto de vista computacional,

iguais àqueles usados para a verossimilhança, e são geralmente mais intensivos e demorados que os algoritmos de estimação matriciais. Há diversos trabalhos na literatura que abordam a construção de perfis, como os de Li (2000), Lin e Zhang (2002), Macaskill (1993), McCullagh e Tibshirani (1993), Murphy e Vaart (1993).

É importante notar que, substituindo d_i^q na equação (26) pela estatística de Pearson generalizada, obtém-se a máxima pseudo-verossimilhança (MPV). Um resumo das funções de variância mais comuns utilizadas para dados de contagem e os métodos de estimação mais adequados para cada uma delas está disponível na Tabela 3.

Tabela 3 - Algumas funções de variância para dados de contagem, organizadas pela família e método de estimação

Família	Var(Y)	$V(\mu)$	Estimação
Poisson	μ	μ	MV
Quase-Poisson	$\phi\mu$	μ	MQV
BN-2	$\mu(1 + k^{-1}\mu)$	$\mu(1 + k^{-1}\mu)$	MV
BN-2	$\mu(1 + k^{-1}\mu)$	1	MPV
Quase	$\phi\mu^\tau$	μ^τ	MQVE+Perfil

Como foi visto, há muitas possibilidades para abordar a superdispersão, entretanto detectar esse fenômeno requer métodos objetivos. Testes de razão de verossimilhança (ver seção 2.5.2) ou testes de Wald podem ser aplicados, mas necessitam o ajuste de pelos menos dois modelos, um de variância nominal e um com superdispersão. Por outro lado, testes escore requerem apenas que o modelo de variância nominal seja ajustado (MOLLA; MUNISWAMY, 2012). Uma série de estatísticas escore foram propostas para MLGs (ver, por exemplo, os trabalhos de Ganio e Schafer (1992), Lambert e Roeder (1995), Smith e Heitjan (1993)), outras exclusivamente para a regressão Poisson (DEAN; LAWLESS, 1989; XIANG; LEE, 2005) e, para MNLGs, Lin e Li (2008) desenvolveram testes escore para a presença de efeitos aleatórios aditivos. Dentre todas essas propostas, destaca-se a estatística escore geral para contagens e proporções desenvolvida por Dean (1992) pois é de fácil implementação, atende aos pressupostos tanto de MLGs quanto de MNLGs e vem sendo extensivamente estudado (YANG; HARDIN; ADDY, 2009a; YANG; HARDIN; ADDY, 2009b; YANG et al., 2007; ZAMANI; ISMAIL, 2013).

Teste Escore de Dean. Para dados de contagem, o modelo inicial, isto é, sob a hipótese nula, tem distribuição Poisson. Nesse caso, a parametrização pelo valor esperado, no lugar da forma canônica, é mais vantajosa, pois elimina a dependência da matriz local do modelo, $\tilde{\mathbf{X}}$, no cálculo da variância da estatística do teste (DEAN, 1992, p. 492). A fórmula geral do teste é escrita como segue:

$$S = \sum_{i=1}^n \{T_i\} \{\Sigma\}^{-1/2} = \sum_{i=1}^n \left\{ \sigma(\mu_i) \left[\frac{(y_i - \mu_i)^2 - y_i}{2\mu_i^2} \right] \right\} \left\{ \sum_{j=1}^n \frac{\sigma^2(\mu_j)}{2\mu_j^2} \right\}^{-1/2}, \quad (27)$$

em que $\sigma(\cdot)$ é a função de $\boldsymbol{\mu}$ usada para gerar a forma da variância total de $\mathbf{Y}$. Assim, se $\sigma(\mu_i) = \mu_i \Rightarrow \text{Var}(Y_i) = \mu_i(1 + \zeta) = \phi\mu_i$ e, para $\sigma(\mu_i) = \mu_i^2 \Rightarrow \text{Var}(Y_i) = \mu_i + \zeta\mu_i^2 = \mu_i(1 + k^{-1}\mu_i)$.

A estatística S na equação (27), que assintoticamente tem distribuição normal padrão, se refere ao teste da hipótese nula $H_0 : \zeta = 0$. Note que, para a hipótese alternativa ($H_1 : \zeta > 0$), foram feitas apenas suposições a respeito da variância total de Y_i , baseando-se em efeitos aleatórios multiplicativos em μ_i , sem especificar uma distribuição completa para a mistura com a Poisson. Uma abordagem da estimação em modelos de mistura sem especificar a distribuição dos efeitos aleatórios pode ser encontrada em Aitkin (1995).

2.4 Modelos lineares generalizados vetoriais

A classe do modelo linear generalizado vetorial (MLGV) é bastante ampla e foi apresentada primeiramente por Yee e Wild (1996), em um trabalho que também incorpora a extensão imediata dos MLGVs, os chamados modelos lineares aditivos vetoriais. A seguir é apresentada a formulação e justificativa desses modelos, com ligeira mudança de notação daquela utilizada no trabalho original.

2.4.1 Suposições e fórmula geral

Sejam $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ vetores de dimensão q , observados em uma amostra aleatória de uma população com distribuição genericamente denotada por:

$$p(\mathbf{Y} = \mathbf{y}; \boldsymbol{\Lambda}, \phi) = \pi(\mathbf{y}; \lambda_1, \lambda_2, \dots, \lambda_m, \phi), \quad (28)$$

para $\pi(\cdot)$ conhecida, em que $\boldsymbol{\Lambda}^\top = [\lambda_1, \lambda_2, \dots, \lambda_m]^\top$ é o vetor de parâmetros de interesse e ϕ um parâmetro de escala. O logaritmo da função de verossimilhança, supondo $\phi = 1$, de um MLGV para as realizações da equação (28) pode ser escrito como:

$$L[\boldsymbol{\eta}(\boldsymbol{\theta}); \mathbf{y}, \mathbf{x}] = \sum_{i=1}^n L_i[\eta_1(\boldsymbol{\theta}_1), \eta_2(\boldsymbol{\theta}_2), \dots, \eta_m(\boldsymbol{\theta}_m)], \quad (29)$$

em que cada $\eta_j(\cdot)$ é o preditor linear do parâmetro λ_j , tal que $j = 1, 2, \dots, m$, satisfazendo $\eta_j = \boldsymbol{\theta}_j^\top \mathbf{x}_{ij}$; $\boldsymbol{\theta}_j$ é o vetor de comprimento p_j de parâmetros do preditor η_j e $\mathbf{x}_{ij}$ é o vetor de covariáveis, de comprimento p_j , para a i -ésima observação.

Assim como no caso dos MLGs, atribui-se uma função ligação, contínua e diferenciável, entre cada parâmetro da distribuição e seu respectivo preditor, de forma que

$g_j(\lambda_j) = \eta_j$. Nos MLGVs, entretanto, Yee² afirma que $\pi(\cdot)$ não está restrita à família exponencial e engloba distribuições multivariadas; as funções de ligação têm o papel de adequar nos reais o espaço paramétrico de $\pi(\cdot)$ e produzir uma melhor aproximação quadrática do logaritmo da função de verossimilhança; e os parâmetros de $\pi(\cdot)$, para os quais se estabelecem os preditores, não são exclusivamente médias (informação pessoal).

2.4.2 Estimação

A estimação em MLGVs, dados pela fórmula geral na equação (29), pode ser realizada utilizando o algoritmo de mínimos quadrados ponderados iterativamente (JØRGENSEN, 1983; GREEN, 1984), que pode ser escrito como:

$$\boldsymbol{\theta}^{(a+1)} = \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{W}_i^{(a)} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{W}_i^{(a)} \mathbf{z}_i^{(a)} \right), \quad (30)$$

em que $\mathbf{X}_i$ é uma matriz bloco-diagonal $m \times \sum_{j=1}^m p_j$, tal que $\mathbf{X}_i = \text{diag}\{\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{im}\}$; $\mathbf{W}_i^{(a)}$ é a matriz de pesos $m \times m$ na a -ésima iteração do algoritmo, cujo (t, k) -ésimo elemento típico é dado por

$$(\mathbf{W}_i)_{t,k} = -\frac{\partial^2 L_i}{\partial \eta_t \partial \eta_k}; \quad (31)$$

e $\mathbf{z}_i^{(a)}$ é o vetor de comprimento m dos preditores modificados na a -ésima iteração, de forma que $\mathbf{z}_i^{(a)} = \mathbf{X}_i \boldsymbol{\theta}^{(a)} + \mathbf{W}_i^{-1(a)} \mathbf{d}_i^{(a)}$, em que $\mathbf{d}_i = \partial L_i / \partial \boldsymbol{\eta}$. Note que a equação (30) faz uso do algoritmo de Newton-Raphson, sendo suficiente, para o caso dos MLGVs, tomar os valores esperados na equação (31) para utilizar o escore de Fisher.

2.4.3 Resíduos, diagnósticos e gráficos

Medidas de leverage. As medidas de *leverage* são baseadas na matriz de projeção ortogonal $\mathbf{H}$, que, para MLGVs, pode ser escrita como

$$\mathbf{H} = \mathbf{U} \mathbf{X} (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{U}^\top,$$

em que $\mathbf{U}$ é a matriz tal que $\mathbf{U}^\top \mathbf{U} = \mathbf{W}$, ou seja, obtida por meio da decomposição de Cholesky de $\mathbf{W}$; $\mathbf{X} = [\mathbf{x}_{i1}^\top, \mathbf{x}_{i2}^\top, \dots, \mathbf{x}_{im}^\top]^\top$ é a matriz estendida do modelo, de dimensões $nm \times m$; e $\mathbf{W} = \text{diag}\{\mathbf{W}_i\}$ é bloco-diagonal de dimensões $nm \times nm$. Note que $\text{tr}\{\mathbf{H}\} = p = \sum_{j=1}^m p_j$.

A diagonal de $\mathbf{H}$, o vetor denotado por $\mathbf{h} = [\mathbf{h}_1^\top \mid \mathbf{h}_2^\top \mid \dots \mid \mathbf{h}_m^\top]^\top$, tem comprimento nm . O gráfico de índices de $\mathbf{h}_j$ pode revelar observações influentes nas estimativas dos parâmetros do j -ésimo preditor.

²YEE, T.W. **VGLM**. Mensagem recebida por <davibg@usp.br> em 29 abr. 2015.

Resíduos do tipo *working*.

$$\mathbf{r}_i^W = \mathbf{z}_i - \hat{\boldsymbol{\eta}}_i = \mathbf{W}_i^{-1} \mathbf{d}_i,$$

de forma que $\mathbf{r}^W$ é uma matriz $n \times m$. As matrizes $\mathbf{W}_i$ e $\mathbf{d}_i$ são computadas na finalização do algoritmo, ou seja, na iteração em que a condição de convergência é satisfeita.

Os resíduos do tipo *working* servem de base para a obtenção de outros tipos de resíduos, mas, a princípio, não são de interesse direto.

Resíduos de Pearson.

$$\mathbf{r}_i^P = \mathbf{W}_i^{1/2} \mathbf{r}_i^W,$$

tal que $X^2 = \sum_{i=1}^n \sum_{j=1}^m (r_{ij}^P)^2$ é a estatística qui-quadrado de Pearson e $\mathbf{W}_i^{1/2}$ é obtida por decomposição espectral.

Os mesmos tipos de gráficos de resíduos usados para MNLGs podem ser construídos para MLGVs; a diferença está no número de gráficos necessários: um para cada preditor. Assim, se se deseja construir, por exemplo, gráficos de resíduos de Pearson *versus* preditores ajustados ($\hat{\boldsymbol{\eta}}$), serão necessários m gráficos um para cada $\hat{\boldsymbol{\eta}}_j$, com a j -ésima coluna de $\mathbf{r}^P$.

Resíduos de Pearson estudentizados. Rearranjando o vetor $\mathbf{h}$, que é a diagonal da matriz de projeção $\mathbf{H}$, como sendo $\mathbf{h} = [\mathbf{h}_1^\top \mid \mathbf{h}_2^\top \mid \cdots \mid \mathbf{h}_n^\top]^\top$, em que cada $\mathbf{h}_i$ tem comprimento m , os resíduos de Pearson estudentizados podem ser escritos como:

$$\mathbf{r}_i^{P*} = \mathbf{r}_i^P \circ (1 - \mathbf{h}_i^\top)^{-1/2},$$

em que $\mathbf{a} \circ \mathbf{b}$ é o produto escalar entre $\mathbf{a}$ e $\mathbf{b}$, para $\mathbf{a}$ e $\mathbf{b}$ de mesmas dimensões.

Componentes de *deviance* e componentes de *deviance* estudentizados. Estes tipos de resíduos apenas estão definidos para os MLGVs quando podem ser escritos como MLGs e as fórmulas para obtenção dessas quantidades são, portanto, iguais àquelas das equações (17) e (18), na seção 2.3.3.

Medidas de influência e curvatura locais. Seguindo as mesmas justificativas, apresentadas na seção 2.3.3, de perturbação caso-a-caso em torno de um vetor $\mathbf{v}_0$ sobre a verossimilhança, as medidas de influência e curvatura locais em MLGVs podem ser obtidas para o conjunto das estimativas $\hat{\boldsymbol{\theta}}$ e para subconjuntos de interesse. Particularmente de interesse, são as estimativas $\hat{\boldsymbol{\theta}}_j$ respectivamente associadas a cada um dos preditores $\boldsymbol{\eta}_j$, para $j = 1, 2, \dots, m$. Isto é obtido simplesmente pelas partições da matriz de informação observada,

$$\mathbf{K} = \sum_{i=1}^n \mathbf{X}_i^\top (\mathbf{d}_i \mathbf{d}_i^\top) \mathbf{X}_i = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \cdots & \mathbf{K}_{1m} \\ \mathbf{K}_{21} & \mathbf{K}_{22} & \cdots & \mathbf{K}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_{m1} & \mathbf{K}_{m2} & \cdots & \mathbf{K}_{mm} \end{bmatrix},$$

e da matriz de derivadas segundas da verossimilhança perturbada $\mathbf{\Delta} = [\mathbf{\Delta}_1^\top | \mathbf{\Delta}_2^\top | \mathbf{\Delta}_3^\top | \cdots | \mathbf{\Delta}_m^\top]^\top$. Assim, obtém-se, para preditores de parâmetros assintoticamente independentes, $\mathbf{C}_{l_{max,j}}$ e $l_{max,j}$ calculando os auto-valores e auto-vetores de $\mathbf{B}_j = \mathbf{\Delta}_j^\top \mathbf{K}_{jj}^{-1} \mathbf{\Delta}_j$ ou, para o caso geral, $\mathbf{B}_j^* = \mathbf{\Delta}^\top \mathbf{K}_j^{*-1} \mathbf{\Delta}$, em que os elementos $\mathbf{K}_j^{*-1}$ contêm os elementos de $\mathbf{K}$ para as partições relacionadas a j e $\mathbf{0}$, em caso contrário. Podem ser considerados também subconjuntos de $\hat{\theta}_j$, seguindo os mesmos passos descritos na seção 2.3.3.

2.5 Seleção de modelos

2.5.1 Teste de Wald

Os testes de Wald são baseados na normalidade assintótica dos estimadores de máxima verossimilhança dos parâmetros e podem ser aplicados tanto em MNLGs quando em MLGVs. Sob condições de regularidade, $\hat{\theta} \sim N_p(\theta, \mathbf{F}^{-1})$, em que N_p denota a distribuição normal p -variada e $\mathbf{F}^{-1}$ é a inversa da matriz de informação esperada, que, para os modelos deste trabalho, é equivalente à matriz de covariâncias assintóticas (ver Cordeiro e Santana (2008) para as matrizes de covariância assintóticas de ordem superior em MNLGs). Invariavelmente para o parâmetros de dispersão/escala ($\phi = 1$), a matriz $\mathbf{F}$ nos MNLGs é dada por $\tilde{\mathbf{X}}^\top \mathbf{W} \tilde{\mathbf{X}}$ e, nos MLGVs, pode ser escrita como $\mathbf{X}^\top \mathbf{W} \mathbf{X}$ (YEE; HASTIE, 2003), seguindo as notações das seções 2.3.2 e 2.4.2, respectivamente.

O t -ésimo elemento na diagonal de $\mathbf{F}^{-1}$, denotado por $\text{Var}(\hat{\theta}_t)$, representa a variância assintótica de $\hat{\theta}_t$. Assim, para testar a hipótese simples $H_0 : \hat{\theta}_t = \theta_{H_0}$ *versus* $H_1 : \hat{\theta}_t \neq \theta_{H_0}$, pode-se utilizar a estatística

$$S_W = \frac{\hat{\theta}_t - \theta_{H_0}}{\sqrt{\text{Var}(\hat{\theta}_t)}},$$

que, sob H_0 , tem distribuição normal padrão. Testes de Wald são aplicados a um ou poucos parâmetros, de forma que, para seleção de modelos, prefere-se, geralmente, utilizar o teste da razão de verossimilhanças ou critérios de informação.

2.5.2 Teste da razão de verossimilhanças

A construção do teste de razão de verossimilhanças é, do ponto de vista prático, simples de interpretar e é de fácil de implementação, tanto para MNLGs quanto para MLGVs: sejam dois modelos aninhados para o mesmo conjunto de dados, ajustados de

forma independente. Seja a verossimilhança do primeiro denotada por L_1^* e, a do segundo, por L_2^* . Se os dois modelos têm o mesmo componente aleatório e parâmetro de dispersão conhecido, mas o número de parâmetros nos preditores são diferentes, tal que o primeiro possua p parâmetros e o segundo q parâmetros, para $p > q$, então a estatística

$$S_{lrt} = -2 \log \frac{L_2^*}{L_1^*} \quad (32)$$

tem assintoticamente distribuição de qui-quadrado com $p - q$ graus de liberdade (GL). Pelo princípio da parcimônia, o modelo cuja verossimilhança é dada por L_2^* é selecionado se S_{lrt} não for significativa a $100(1 - \alpha)\%$.

A razão de verossimilhanças pode ser usada também como um teste de bondade de ajuste do modelo, substituindo a quantidade no denominador da equação (32) pela verossimilhança do modelo saturado, com um parâmetro para cada observação. Neste caso, Cordeiro e Paula (1989) propuseram um teste melhorado, baseado na esperança de S_{lrt} .

Apesar de muito úteis em contextos específicos, os testes de razão de verossimilhança não podem ser aplicados para a seleção de preditores não-lineares não aninhados, uma situação frequente se forem considerados os modelos tradicionais do TA (seção 2.2); tampouco para decidir sobre a escolha do componente aleatório. Para tal, os critérios de informação devem ser empregados.

2.5.3 Critérios de Informação

Os critérios de informação estão fundamentados no trabalho de Kullback e Leibler (1951) que propuseram uma fórmula para medir a perda de informação (medida de K-L) decorrente do ajuste de um modelo aproximado, em detrimento do “modelo verdadeiro”. Neste contexto, o modelo selecionado é aquele que implica na menor perda de informação. Entretanto, a realidade completa é desconhecida, de forma que a medida de K-L não pode ser diretamente aplicada. Para contornar esse problema, pode-se tomar o valor esperado da informação, expressa como

$$\begin{aligned} E[I_{KL}(f, g)] &= E \left[\int f(x) \log \left(\frac{f(x)}{g(x; \boldsymbol{\lambda})} \right) dx \right] \\ &= E_f \{ \log [f(x)] \} - E_f \{ \log [g(x; \boldsymbol{\lambda})] \}, \end{aligned}$$

em que, seguindo a notação de Burnham e Andersen (2004), $I_{KL}(f, g)$ é a medida da perda de informação de K-L; $f(\cdot)$ representa a realidade completa e desconhecida; $g(\cdot)$ é o modelo probabilístico proposto para aproximar f ; $\boldsymbol{\lambda}$ são os parâmetros da distribuição g ; e E_f indica que os valores esperados devem ser tomados em relação a f .

Note que $E_f \{ \log [f(x)] \}$ é uma constante, que independe do modelo aproximado g

e dos dados observados. Assim, apenas a quantidade $E_f \{ \log [g(x; \boldsymbol{\lambda})] \}$ precisa ser obtida para servir como critério de seleção entre os possíveis (muitos) modelos de aproximação, sob consideração para ajuste a um conjunto particular de dados.

A estratégia adotada por Akaike (1992, p. 611) para obter uma estimativa da medida de informação de K-L foi primeiro procurar, seguindo a notação de Burnham e Andersen (2004), uma estimativa da dupla esperança

$$E_{\mathbf{y}} \left\{ E_x \left\{ \log \left[g(x; \hat{\boldsymbol{\lambda}}, \mathbf{y}) \right] \right\} \right\}, \quad (33)$$

baseada nas estimativas de máxima verossimilhança $\hat{\boldsymbol{\lambda}}$, obtidas pelo modelo g a partir dos dados observados $\mathbf{y}$; e na suposição de que x e $\mathbf{y}$ são independentes entre si, mas provenientes do mesmo modelo proposto g . Fazendo uso de argumentos assintóticos, Akaike (1973 apud BURNHAM; ANDERSEN, 2004) mostrou que a verossimilhança em $\hat{\boldsymbol{\lambda}}$ é um estimador viesado para a dupla esperança na equação (33), por um termo aditivo aproximadamente igual ao número de parâmetros no modelo, p . Assim, o estimador assintoticamente não-viesado pode ser escrito como:

$$-p + \sum_{i=1}^n L_i(\hat{\boldsymbol{\lambda}}; \mathbf{y}) = \hat{E} [I_{\text{KL}}(f, \hat{g})], \quad (34)$$

em que $L_i(\cdot)$ é o logaritmo da função de verossimilhança na i -ésima observação; e $\hat{g} = g(\cdot; \hat{\boldsymbol{\lambda}})$. A partir do resultado na equação (34), o critério de informação de Akaike (AIC, na sigla em inglês) é obtido apenas pela multiplicação por uma constante -2 :

$$\text{AIC} = -2 \sum_{i=1}^n L_i(\hat{\boldsymbol{\lambda}}; \mathbf{y}) + 2p. \quad (35)$$

Existem conexões entre as propriedades da razão de verossimilhanças, o critério de informação de Shannon-Wiener (SHANNON, 1948), o conceito de informação de Fisher e as medidas de entropia na termodinâmica (TSALLIS, 1988), que, segundo deLeeuw (1992), justificam a multiplicação por -2 na equação (35). Além disso, esta formulação do AIC é válida para $n \rightarrow \infty$, de forma que, na maioria dos problemas práticos, deve-se utilizar a versão para pequenas amostras, como segue:

$$\text{AIC}_c = -2 \sum_{i=1}^n L_i(\hat{\boldsymbol{\lambda}}; \mathbf{y}) + 2p + \frac{2p(p+1)}{n-p-1},$$

que é corrigida para viés de segunda ordem (HURVICH; TSAI, 1989; HURVICH; TSAI, 1991; HURVICH; TSAI, 1995) e é recomendada quando $n/p < 40$ (BURNHAM; ANDERSEN, 2004, p. 270).

Dada essa dependência intrínseca que os valores de AIC e de AIC_c têm de constantes relacionadas ao tamanho da amostra, é imprescindível definir uma escala irrestrita para a

comparação dos modelos candidatos. A simples diferença do valor de AIC de cada modelo ajustado e do mínimo entre os valores dos AICs de todos os modelos ajustados, chamada de “AIC escalonado”, é uma escala que permite a comparação direta entre os modelos candidatos e é escrita como:

$$\Delta_j^{(\text{AIC})} = \text{AIC}_j - \min \{ \text{AIC}_1, \text{AIC}_2, \dots, \text{AIC}_r \},$$

em que AIC_j é o valor de AIC para o j -ésimo modelo, tal que $j = 1, 2, \dots, r$. A medida $\Delta_j^{(\text{AIC})}$ impõe que o melhor ajuste entre os r candidatos seja igual a 0. Na Tabela 4 podem ser encontrados os valores $\Delta_j^{(\text{AIC})}$ e sua respectiva força de evidência empírica (FEE), que pode ser usada como referência para decidir entre os modelos.

Outra escala de referência baseada nos valores de AIC é chamada “pesos de Akaike”, definidos como:

$$\omega_j^{(\text{AIC})} = \frac{\exp \left\{ -\Delta_j^{(\text{AIC})} / 2 \right\}}{\sum_{j=1}^r \exp \left\{ -\Delta_j^{(\text{AIC})} / 2 \right\}}, \quad (36)$$

em que $\omega_j^{(\text{AIC})}$ pode ser interpretado como a probabilidade de que o j -ésimo modelo seja a melhor aproximação da “realidade”, isto é, tenha a menor perda na informação de K-L.

Tabela 4 - Valores de AIC escalonado e sua respectiva força de evidência, considerando um conjunto arbitrário de r modelos ajustados ao mesmo conjunto de dados. $\Delta_j^{(\text{AIC})}$: AIC escalonado do j -ésimo modelo; FEE: força de evidência empírica à favor do j -ésimo modelo. Modificado de Burnham e Andersen (2002, p. 70)

$\Delta_j^{(\text{AIC})}$	FEE
0–2	Muito forte
2–4	Forte
4–7	Razoável
7–10	Fraca
>10	Muito fraca

Existem, além do AIC, outros critérios baseados na medida de K-L, como o TIC (TAKEUCHI, 1976 apud SHIBATA, 1997) e o BIC (SCHWARZ, 1978). Há um contexto diferente dentro do qual um ou outro critério apresenta melhor performance. Para pequenas amostras e na presença de múltiplos efeitos, pequenos e não-estimáveis, o uso do AIC é recomendado (BURNHAM; ANDERSEN, 2002, p. 285) – uma situação típica em experimentos de microbiologia.

O TA não é uma exceção à regra das pequenas amostras e das pequenas variações nas unidades experimentais dos experimentos em microbiologia e, como já visto, é comum a presença de superdispersão. Para modelos de quase-verossimilhança, visando dados de contagem superdispersos, há uma adaptação do próprio AIC, o quase-AIC ou QAIC:

$$\text{QAIC} = \frac{-2 \sum_{i=1}^n L_i(\hat{\boldsymbol{\lambda}}; \mathbf{y})}{\hat{c}} + 2p,$$

em que $\hat{c}$ é o parâmetro de dispersão estimado pela equação (25) (a estatística generalizada de Pearson) apenas no modelo com maior número de parâmetros, e os demais modelos são sub-modelos aninhados (BURNHAM; ANDERSEN, 2002, p. 68). A versão corrigida para viés de segunda ordem é simplesmente:

$$\text{QAIC}_c = \text{QAIC} + \frac{2p(p+1)}{n-p-1}.$$

Note que, quando são utilizadas MQV, MQVE ou MPV, nem QAIC e nem QAIC_c servem para comparar modelos não-aninhados ou diferentes funções de variância, limitando bastante seu uso. Assim, a seleção nesses casos, precisa ser baseada apenas em análise de resíduos e diagnósticos.

3 MATERIAL

Os dados das Tabelas 5 e 6 se referem, respectivamente, aos resultados de experimentos desenvolvidos com as linhagens TA98 e TA100 de *S. typhimurium* no Laboratório de Mutagênese, Departamento de Biologia, Instituto de Biociências, UNESP/Rio Claro-SP e foram realizados em parceria com a Dra. Thaís C. C. Fernandes, sob supervisão da Profa. Dra. Maria Aparecida Marin-Morales.

O delineamento experimental foi do tipo inteiramente casualizado em esquema fatorial para os tratamentos. Foram avaliadas diversas concentrações de dois inseticidas, além dos controles negativo (concentração zero, apenas água estéril) e positivo, seguindo as recomendações para elaboração do ensaio de incorporação em placas e pré-incubação de *Salmonella*/microsoma disponíveis em Umbuzeiro et al. (2004).

Em conformidade com o referido protocolo nacional, os experimentos foram realizados na presença (+) e na ausência (−) do sistema de ativação metabólica S9, utilizando, para o controle positivo, a substância 2-aminoantraceno para S9(+) nos experimentos com ambas as linhagens, e as substâncias 4-nitroquinolina e azida sódica para S9(−), respectivamente para as linhagens TA98 e TA100. Os dados obtidos nas placas tratadas com os controles positivos não são de interesse estatístico e, portanto, foram omitidos neste trabalho. Apenas os resultados das placas que não apresentaram sinais de contaminação externa foram considerados para análise.

Os inseticidas utilizados nos experimentos foram o Fipronil e o Tiametoxam, que se referem respectivamente aos produtos comerciais Standak® (250 g/L de princípio ativo) e Cruiser® (350 g/L de princípio ativo). Estes agroquímicos são amplamente aplicados nas lavouras do Brasil e têm sido alvo de estudos e críticas nas mais diversas áreas da ciência nacional e internacional (ver, por exemplo, os trabalhos de Gibbons, Morrissey e Mineau (2015), Pedro (2008), Pereira (2010), Rosa (2014), Sluijs et al. (2015)). As quantidades recomendadas para campo são de 25 g/L de Fipronil e de 3 g/L de Tiametoxam, de forma que foram essas as doses de referência utilizadas nos experimentos.

Tabela 5 - Número observado de colônias revertentes no experimento utilizando o protocolo padrão do teste de Ames (*Salmonella*/microsoma), versão de incorporação em placas, com a linhagem TA98 de *S. typhimurium*, em diferentes concentrações dos inseticidas Tiametoxam e Fipronil, na ausência e na presença da fração microsomal S9 (Sistema de ativação metabólica)

Produto	Sistema	Concentração (g/L)	Número de colônias revertentes							
Tiametoxam	S9(+)	3	24	30	35					
Tiametoxam	S9(+)	$3 \cdot 2^{-1}$	29	35	37					
Tiametoxam	S9(+)	$3 \cdot 2^{-2}$	19	22	24					
Tiametoxam	S9(+)	$3 \cdot 2^{-3}$	16	19	17					
Tiametoxam	S9(+)	$3 \cdot 2^{-4}$	16	24	23					
Tiametoxam	S9(+)	$3 \cdot 2^{-5}$	15	19	10					
Tiametoxam	S9(+)	$3 \cdot 2^{-6}$	23	18	12					
Tiametoxam	S9(+)	$3 \cdot 2^{-7}$	15	16	18					
Tiametoxam	S9(+)	$3 \cdot 2^{-8}$	31	41	30					
Tiametoxam	S9(+)	$3 \cdot 2^{-9}$	30	29	30					
Tiametoxam	S9(+)	$3 \cdot 2^{-10}$	32	22	32					
Tiametoxam	S9(+)	0	29	22	23	25	16	33	32	32
Tiametoxam	S9(-)	3	38	42	28					
Tiametoxam	S9(-)	$3 \cdot 2^{-1}$	34	31	30					
Tiametoxam	S9(-)	$3 \cdot 2^{-2}$	17	25						
Tiametoxam	S9(-)	$3 \cdot 2^{-3}$	23	27	24					
Tiametoxam	S9(-)	$3 \cdot 2^{-4}$	27	41	29					
Tiametoxam	S9(-)	$3 \cdot 2^{-5}$	23	28	25					
Tiametoxam	S9(-)	$3 \cdot 2^{-6}$	21	32	24					
Tiametoxam	S9(-)	$3 \cdot 2^{-7}$	24	26	22					
Tiametoxam	S9(-)	$3 \cdot 2^{-8}$	29	38	30					
Tiametoxam	S9(-)	$3 \cdot 2^{-9}$	35	37	45					
Tiametoxam	S9(-)	0	18	27	31	26	31	26	36	21
Fipronil	S9(+)	25	526	502	583					
Fipronil	S9(+)	$25 \cdot 10^{-1}$	57	64	53					
Fipronil	S9(+)	$25 \cdot 10^{-2}$	34	31	37					
Fipronil	S9(+)	$25 \cdot 10^{-3}$	39	32	26					
Fipronil	S9(+)	$25 \cdot 10^{-4}$	28	34	32					
Fipronil	S9(+)	$25 \cdot 10^{-5}$	25	26	27					
Fipronil	S9(+)	0	28	25	33	35	32			
Fipronil	S9(-)	25	623	650	548					
Fipronil	S9(-)	$25 \cdot 10^{-1}$	112	75	88					
Fipronil	S9(-)	$25 \cdot 10^{-2}$	42	45	38					
Fipronil	S9(-)	$25 \cdot 10^{-3}$	30	43	49					
Fipronil	S9(-)	$25 \cdot 10^{-4}$	36	48	37					
Fipronil	S9(-)	$25 \cdot 10^{-5}$	39	35	46					
Fipronil	S9(-)	0	30	24	33	24	32			

Tabela 6 - Número observado de colônias revertentes no experimento utilizando o protocolo padrão do teste de Ames (*Salmonella*/microsoma), versão de incorporação em placas, com a linhagem TA100 de *S. typhimurium*, em diferentes concentrações dos inseticidas Tiametoxam e Fipronil, na ausência e na presença da fração microsomal S9 (Sistema de ativação metabólica)

Produto	Sistema	Concentração (g/L)	Número de colônias revertentes						
Tiametoxam	S9(+)	3	105	99	110				
Tiametoxam	S9(+)	$3 \cdot 2^{-1}$	109	115	123				
Tiametoxam	S9(+)	$3 \cdot 2^{-2}$	248	250	227				
Tiametoxam	S9(+)	$3 \cdot 2^{-3}$	156	140	160				
Tiametoxam	S9(+)	$3 \cdot 2^{-4}$	139	146	134				
Tiametoxam	S9(+)	$3 \cdot 2^{-5}$	123	133	128				
Tiametoxam	S9(+)	$3 \cdot 2^{-6}$	133	132	129				
Tiametoxam	S9(+)	$3 \cdot 2^{-7}$	153	149	273				
Tiametoxam	S9(+)	$3 \cdot 2^{-8}$	117	124	102				
Tiametoxam	S9(+)	$3 \cdot 2^{-9}$	123	113	113				
Tiametoxam	S9(+)	$3 \cdot 2^{-10}$	96	111	139				
Tiametoxam	S9(+)	0	122	131	144	141	141	111	105 103
Tiametoxam	S9(-)	3	89	56					
Tiametoxam	S9(-)	$3 \cdot 2^{-1}$	87	86					
Tiametoxam	S9(-)	$3 \cdot 2^{-2}$	294	212	249				
Tiametoxam	S9(-)	$3 \cdot 2^{-3}$	127	152	138				
Tiametoxam	S9(-)	$3 \cdot 2^{-4}$	142	150	142				
Tiametoxam	S9(-)	$3 \cdot 2^{-5}$	135	161	158				
Tiametoxam	S9(-)	$3 \cdot 2^{-6}$	197	164	173				
Tiametoxam	S9(-)	$3 \cdot 2^{-7}$	154	149	140				
Tiametoxam	S9(-)	$3 \cdot 2^{-8}$	75	41					
Tiametoxam	S9(-)	$3 \cdot 2^{-9}$	64	62					
Tiametoxam	S9(-)	$3 \cdot 2^{-10}$	42	84					
Tiametoxam	S9(-)	0	83	134	143	135	148	79	71
Fipronil	S9(+)	250	756	1430	1020				
Fipronil	S9(+)	$250 \cdot 2^{-1}$	960	1040	153				
Fipronil	S9(+)	$250 \cdot 10^{-2}$	138	140	144				
Fipronil	S9(+)	$250 \cdot 10^{-3}$	167	154	153				
Fipronil	S9(+)	$250 \cdot 10^{-4}$	149	149	157				
Fipronil	S9(+)	$250 \cdot 10^{-5}$	106	149	168				
Fipronil	S9(+)	$250 \cdot 10^{-6}$	133	136	129				
Fipronil	S9(+)	$250 \cdot 10^{-7}$	129	126	109				
Fipronil	S9(+)	$250 \cdot 10^{-8}$	116	124	117				
Fipronil	S9(+)	$250 \cdot 10^{-9}$	121	123	120				
Fipronil	S9(+)	0	155	180	159	174	191	100	119 111
Fipronil	S9(-)	250	994	1476	1248				
Fipronil	S9(-)	$250 \cdot 2^{-1}$	500	756	678				
Fipronil	S9(-)	$250 \cdot 10^{-1}$	166	188	183				
Fipronil	S9(-)	$250 \cdot 10^{-2}$	169	166	176				
Fipronil	S9(-)	$250 \cdot 10^{-3}$	186	178	193				
Fipronil	S9(-)	$250 \cdot 10^{-4}$	183	171	170				
Fipronil	S9(-)	$250 \cdot 10^{-5}$	172	165	156				
Fipronil	S9(-)	$250 \cdot 10^{-6}$	161	199	158				
Fipronil	S9(-)	$250 \cdot 10^{-9}$	111	103	143				
Fipronil	S9(-)	0	199	192	143	202	164	130	84 205

4 METODOLOGIA

4.1 Modelos tradicionais

4.1.1 Estudos de simulação

A realização de simulações dos modelos se faz importante, particularmente no contexto dos MNLGs, para que possam ser avaliados tanto o nível de não-linearidade do modelo quanto os comportamentos assintóticos dos estimadores, além de possuir a vantagem de que as possíveis reparametrizações sejam prontamente obtidas e analisadas (RATKOWSKY, 1983; RATKOWSKY, 1989). Os estudos de simulação são mais práticos do que obter as medidas de curvatura apresentadas, para respostas estritamente com distribuição normal, por Bates e Watts (1980), Bates e Watts (1988) e, para distribuições pertencentes à família exponencial, por (WEI, 1998, p. 39-46).

Neste trabalho, foram adotadas simulações de Monte Carlo, seguindo o procedimento:

1. Escolher a forma da média verdadeira (preditores tradicionais);
2. Sortear os valores dos parâmetros;
3. Escolher a distribuição (componente aleatório);
4. Ajustar o modelo adequado (ao preditor da média, componente aleatório e função de ligação logarítmica);
5. Repetir os passos 3 e 4 1000 vezes até contemplar todas as distribuições;
6. Repetir os passos 1 a 4 100 vezes até contemplar todos os preditores.

Este procedimento permitiu a avaliação de diferentes cenários, dados pela combinação de (preditor $\times$ vetor de parâmetros $\times$ componente aleatório $\times$ nível de superdispersão).

As formas escolhidas para as médias verdadeiras (Passo 1) foram aquelas dos modelos tradicionais (equações (5) a (10)), cuja avaliação a respeito do nível de não-linearidade e do comportamento assintótico foi o objetivo deste estudo de simulação. A variável explicativa presente em todos os preditores que, neste contexto, representa as doses de exposição, foi composta por 6 níveis com três repetições cada, no mesmo esquema proposto em Umbuzeiro et al. (2004), o protocolo nacional para o TA.

A variação nos valores dos parâmetros (Passo 2) de cada preditor é de importância, pois por si só podem comprometer a qualidade do ajuste, ou porque estão muito próximos dos limites do espaço paramétrico ou porque refletem em valores esperados que dificultam a convergência do algoritmo de estimação. Os 100 vetores diferentes para os valores dos

parâmetros de cada preditor foram obtidos por realizações de distribuições uniformes, cujos máximos e mínimos podem ser visualizados na Tabela 7. Estes máximos e mínimos foram escolhidos tais que, dados os níveis da variável explicativa, produzissem valores esperados realistas do ponto de vista biológico.

Tabela 7 - Máximos e mínimos dos parâmetros, uniformemente distribuídos, utilizados para o estudo de simulação de cada modelo das equações (5) a (10). A quantidade p se refere ao número de parâmetros no respectivo preditor e $[min; max]$ ao intervalo considerado para a distribuição uniforme

Preditor	p	β_0		β_1		β_2		γ	
		min	max	min	max	min	max	min	max
Bernstein	2	4,317	5,298	0	1	-	-	-	-
Breslow	3	4,317	5,298	2	2,3	-	-	0,5	1,7
Krewski	3	4,317	5,298	1,2	2,6	-	-	0,03	0,08
Margolin	3	0,267	0,339	0,3	0,9	-	-	0,25	0,7
Myers	3	4,317	5,298	2	3,1	-	-	0,2	0,3
Stead	4	4,317	5,298	2	3,5	1	3	0,3	0,45

Neste estudo foram consideradas as distribuições Poisson, BN-2 e PG (Passo 3). O objetivo em realizar simulações, tanto da BN-2 quanto da PG, consiste em contemplar a possível presença de superdispersão, um fenômeno reportado por diversos autores que analisaram dados provenientes do TA. O nível de superdispersão aumenta à medida que o parâmetro k da BN-2 diminui ou à medida que o parâmetro ϕ da Poisson generalizada aumenta. Foram considerados 3 valores diferentes para k (100, 50 e 25) e 3 para ϕ (2, 4 e 8).

Foram implementadas no R e utilizadas as funções `rames()` para gerar as contagens simuladas (seção 4.5.8) e `fit.ames()` para o ajuste dos modelos (seção 4.5.1). Seguindo as recomendações de Ratkowsky (1983) e de Ratkowsky (1989), foram obtidas, em cada cenário, as estatísticas de teste:

(i) para o viés,

$$S_b = \frac{b_{tj}(\hat{\theta}_{tj}, \theta_{tj})\sqrt{N}}{\sqrt{\text{Var}_{tj}(\theta_{tj})}}; \quad (37)$$

(ii) para excesso de variância,

$$S_v = (N - 1) \frac{m_{tj,2}}{\text{Var}_{tj}(\theta_{tj})}; \quad (38)$$

(iii) para a simetria,

$$S_{g_1} = g_{1tj} = \frac{m_{tj,3}}{\sqrt{m_{tj,2}^3}}; \quad (39)$$

(iv) para a curtose,

$$S_{g_2} = g_{2tj} = \frac{m_{tj,4}}{\sqrt{m_{tj,2}^2}} - 3. \quad (40)$$

Nas equações (37) a (40), os índices i e t se referem à i -ésima simulação ($i = 1, 2, \dots, N$; $N = 1000$) com o t -ésimo vetor de parâmetros ($t = 1, 2, \dots, 100$). Assim, dentro do vetor de parâmetros gerado conforme a Tabela 7, são feitos os cálculos do viés, dos momentos amostrais e dos coeficientes de curtose e assimetria para o j -ésimo parâmetro e sua estimativa correspondente ($j = 1, 2, \dots, p$). Note que o viés $b_{tj}(\hat{\theta}_{tj}, \theta_{tj}) = \tilde{\theta}_{tj} - \theta_{tj}$, tal que $\tilde{\theta}_{tj} = \sum_{i=1}^N \hat{\theta}_{itj}/N$. Também são importantes a quantidade $\text{Var}_{tj}(\theta_{tj})$, que é o j -ésimo elemento da diagonal da matriz de covariâncias assintóticas; e os k -ésimos momentos amostrais em torno da média, $m_{tj,k}$, dados por $\sum_{i=1}^N (\hat{\theta}_{itj} - \tilde{\theta}_{tj})^k / N$.

Ratkowsky (1983) argumenta que a estatística S_b , da equação (37), tem distribuição normal padrão e que a estatística S_v , da equação (38), tem distribuição de qui-quadrado com $N - 1$ graus de liberdade. As estatísticas S_{g_1} , da equação (39), e S_{g_2} , da equação (40), têm distribuições normais com média 0 e variâncias dadas respectivamente por $6/N$ e $24/N$ (SNEDECOR; COCHRAN, 1980). Todos os testes de hipótese realizados nos estudos de simulação deste trabalho foram feitos a 95% de significância.

Também seguindo as recomendações de Ratkowsky (1983) os valores iniciais dos parâmetros para o ajuste do modelo, com o intuito de assegurar a convergência do algoritmo, são idênticos aos valores reais, utilizados para realizar as simulações.

4.1.2 Obtenção dos valores iniciais

Diferentemente dos estudos de simulação, os valores reais dos parâmetros são desconhecidos e métodos para a obtenção de valores iniciais para o algoritmo de estimação são necessários.

Utilizando uma adaptação do método de momentos para um sistema de equações, é possível obter boas estimativas iniciais para os parâmetros dos modelos apresentados. Quando se está interessado em estimar uma curva de valores esperados, como é o caso da regressão em MNLGs, em modelos com até três parâmetros, esse procedimento é de aplicação relativamente simples e demanda apenas a solução numérica de uma única equação não-linear. Existem potencialmente duas ou mais soluções reais para o sistema, mas, geralmente, apenas uma está dentro do espaço paramétrico.

É claro que o preditor de cada modelo implica em um sistema diferente de equações, de forma que, a seguir, são apresentados os sistemas para cada modelo um dos modelos tradicionais. Para os modelos cujos preditores consistem de funções lineares ou linearizáveis, nenhum método de obtenção de valor inicial foi aplicado, como é o caso dos modelos de Bernstein (equação (8)) e de Breslow (equação (9)).

Os métodos para obtenção dos valores iniciais foram implementados em R, na função

`iv.ames()` (ver seção 4.5).

Sistema de equações para o modelo de Margolin Para simplicidade de notação, nesta seção e nas seções subsequentes, as quantidades β_0 e β_1 para o modelo de Margolin se referem, respectivamente, aos parâmetros β_0^* e β_1^* da equação (5).

$$\begin{cases} m(1 - e^{-\hat{\beta}_0^{(0)}}) & = \bar{\omega}_0 \\ m \left[1 - e^{-(\hat{\beta}_0^{(0)} + \hat{\beta}_1^{(0)} x_1)} \right] e^{-\hat{\gamma}^{(0)} x_1} & = \bar{\omega}_1 \\ m \left[1 - e^{-(\hat{\beta}_0^{(0)} + \hat{\beta}_1^{(0)} x_2)} \right] e^{-\hat{\gamma}^{(0)} x_2} & = \bar{\omega}_2 \end{cases}$$

em que $\bar{\omega}_k = n_k^{-1} \sum_{i=1}^{n_k} g(y_i; x_i = x_k)$ é a média do número de revertentes observados na k -ésima dose transformados para a escala da função de ligação $g(\cdot)$, tal que $k = \{0, 1, 2\}$ e que $x_0 = 0$. Do sistema, obtém-se diretamente a solução $\hat{\beta}_0^{(0)} = \log(m) - \log(m - \bar{\omega}_0)$, embora não existam soluções analíticas para $\hat{\beta}_1^{(0)}$ e $\hat{\gamma}^{(0)}$. Pode-se, entretanto, simplificar o sistema de modo a obter:

$$\left\{ \hat{\beta}_1^{(0)} = \frac{1}{x_1 x_2} \left[\log \left(\frac{m - \bar{\omega}_0}{m \exp(\frac{\mathbf{z}^* x_1}{x_2}) - \bar{\omega}_1} \right) x_2 + \mathbf{z}^* x_1 \right], \hat{\gamma}^{(0)} = \frac{\mathbf{z}^*}{x_2} \right\} \quad (41)$$

em que $\mathbf{z}^*$ é o vetor de soluções, em relação a z , da equação:

$$z + \log \left(\frac{\bar{\omega}_2 - m}{\bar{\omega}_2 - m} \right) + \frac{x_2}{x_1} \log \left[\frac{m - \bar{\omega}_0}{m \exp(\frac{zx_2}{x_1}) - \bar{\omega}_1} \right] = 0 \quad (42)$$

Utilizando métodos numéricos, como os disponíveis no pacote “rootSolve” de Soetaert (2009), para encontrar $\mathbf{z}^*$ da equação (42) e, então, substituindo em (41), são obtidos os valores iniciais para todos os parâmetros do modelo em questão.

Sistema de equações para o modelo de Myers

$$\begin{cases} \hat{\beta}_0^{(0)} & = \bar{\omega}_0 \\ (\hat{\beta}_0^{(0)} + \hat{\beta}_1^{(0)} x_1) e^{-\hat{\gamma}^{(0)} x_1} & = \bar{\omega}_1 \\ (\hat{\beta}_0^{(0)} + \hat{\beta}_1^{(0)} x_2) e^{-\hat{\gamma}^{(0)} x_2} & = \bar{\omega}_2 \end{cases}$$

Excetuando-se $\hat{\beta}_0^{(0)}$, para o qual a solução é imediata, novamente não existem soluções analíticas do sistema para $\hat{\beta}_1^{(0)}$ e $\hat{\gamma}^{(0)}$. Pode-se, entretanto, reescrever o sistema como:

$$\left\{ \hat{\beta}_1^{(0)} = \frac{\bar{\omega}_1 - \bar{\omega}_0 \exp\left(\frac{\mathbf{z}^* x_1}{x_2}\right)}{x_1 \exp\left(\frac{\mathbf{z}^* x_1}{x_2}\right)}, \hat{\gamma}^{(0)} = \frac{\mathbf{z}^*}{x_2} \right\} \quad (43)$$

em que $\mathbf{z}^*$ é o vetor de soluções, em relação a z , da equação:

$$\bar{\omega}_0 x_1 e^{z + \frac{z x_1}{x_2}} - \bar{\omega}_0 x_2 e^{z + \frac{z x_1}{x_2}} + \bar{\omega}_1 x_2 e^z - \bar{\omega}_2 x_1 x_2 e^{z + \frac{z x_1}{x_2}} = 0 \quad (44)$$

Utilizando métodos numéricos similares para encontrar $\mathbf{z}^*$ em (44) e, então, substituindo em (43), são obtidos os valores iniciais para os parâmetros restantes.

Sistema de equações para o modelo de Stead

$$\begin{cases} \hat{\beta}_0^{(0)} &= \bar{\omega}_0 \\ (\hat{\beta}_0^{(0)} + \hat{\beta}_1^{(0)} x_1^{\beta_2}) e^{-\hat{\gamma}^{(0)} x_1} &= \bar{\omega}_1 \\ (\hat{\beta}_0^{(0)} + \hat{\beta}_1^{(0)} x_2^{\beta_2}) e^{-\hat{\gamma}^{(0)} x_2} &= \bar{\omega}_2 \end{cases}$$

É claro que a solução é imediata para $\hat{\beta}_0^{(0)}$. Mesmo sob a suposição de β_2 conhecido, não existem soluções analíticas do sistema para $\hat{\beta}_1^{(0)}$ e $\hat{\gamma}^{(0)}$. Pode-se, entretanto, reescrever o sistema de modo a obter:

$$\left\{ \hat{\beta}_1^{(0)} = x_1^{-\beta_2} \left(\bar{\omega}_1 e^{-\frac{\mathbf{z}^* x_1}{x_2}} - \bar{\omega}_0 \right), \hat{\gamma}^{(0)} = \frac{\mathbf{z}^*}{x_2} \right\} \quad (45)$$

em que $\mathbf{z}^*$ é o vetor de soluções, em relação a z , da equação:

$$\bar{\omega}_0 (x_2^{\beta_2} - x_1^{\beta_2}) e^{\frac{z(x_1 + x_2)}{x_2}} + \bar{\omega}_2 x_1^{\beta_2} e^{\frac{z x_1}{x_2}} - \bar{\omega}_1 x_2^{\beta_2} e^z = 0 \quad (46)$$

Como, na realidade, β_2 é desconhecido e se está igualmente interessado em obter um valor inicial para a estimativa do parâmetro, o método consiste em atribuir um intervalo de valores ao qual β_2 possa pertencer. Para a ligação logarítmica, β_2 é muito raramente maior que 5, então é suficiente obter todas as soluções da equação (46) para o intervalo discretizado $[1, 5]$. Neste trabalho, esse intervalo foi discretizado em pontos equidistantes, adotando a diferença absoluta de 0,1 entre quaisquer dois pontos subsequentes, de forma que os valores atribuídos para potenciais $\hat{\beta}_2^{(0)}$ foram $\{1,0; 1,1; 1,2; \dots; 4,8; 4,9; 5,0\}$.

Nem todos os valores potenciais para $\hat{\beta}_2^{(0)}$ geram soluções reais na equação (46), dependendo dos valores de $\bar{\omega}_0$, $\bar{\omega}_1$, $\bar{\omega}_2$, x_1 e x_2 e, como β_2 é geralmente pequeno, isto é, pouco maior que 1, escolheu-se o menor valor de $\hat{\beta}_2^{(0)}$ que, além de gerar soluções reais em (46), produziam valores para $\hat{\beta}_1^{(0)}$ e $\hat{\gamma}^{(0)}$ dentro do espaço paramétrico de β_1 e γ quando

substituídas na equação (45).

Sistema de equações para o modelo de Krewski

$$\begin{cases} \hat{\beta}_0^{(0)} &= \bar{\omega}_0 \\ (\hat{\beta}_0^{(0)} + \hat{\beta}_1^{(0)} x_1) e^{-\hat{\gamma}^{(0)} x_1^\alpha} &= \bar{\omega}_1 \\ (\hat{\beta}_0^{(0)} + \hat{\beta}_1^{(0)} x_2) e^{-\hat{\gamma}^{(0)} x_2^\alpha} &= \bar{\omega}_2 \end{cases}$$

A solução para $\hat{\beta}_0^{(0)}$ é obtida diretamente. Entretanto, soluções numéricas para $\hat{\beta}_1^{(0)}$ e $\hat{\gamma}^{(0)}$ podem ser obtidas para o seguinte sistema simplificado:

$$\left\{ \hat{\beta}_1^{(0)} = \frac{\bar{\omega}_2 \exp \left[-\mathbf{z}^* \left(\frac{x_2}{x_1} \right)^\alpha \right] - \bar{\omega}_0}{x_2}, \hat{\gamma}^{(0)} = \mathbf{z}^* \right\} \quad (47)$$

em que $\mathbf{z}^*$ é o vetor de soluções, em relação a z , da equação:

$$\bar{\omega}_0(x_1 - x_2) \exp \left\{ z \left[1 + \left(\frac{x_2}{x_1} \right)^\alpha \right] \right\} + \bar{\omega}_1 x_2 \exp \left[z \left(\frac{x_1}{x_2} \right)^\alpha \right] - \bar{\omega}_2 x_1 e^z = 0 \quad (48)$$

Utilizando métodos numéricos, como a função `uniroot.all()`, disponível em Soetaert (2009), para encontrar $\mathbf{z}^*$ em (48) e, então, substituindo em (47), são obtidos os valores iniciais para todos os parâmetros do modelo em questão.

Valores iniciais obtidos por ajuste de submodelos. Além do método do sistema de equações, é possível obter valores iniciais para os parâmetros de modelos não-lineares por meio do ajuste de modelos mais simples, lineares ou não-lineares. A vantagem desse tipo de abordagem é que toda a informação disponível nos dados é utilizada, ainda que, em determinadas situações, seja difícil definir um submodelo simples a partir de um preditor não-linear complexo.

Particularmente para os modelos do TA, é possível subdividir a curva de resposta em duas partes: para doses suficientemente baixas, ajusta-se apenas a parte do preditor associado à mutagenicidade e, a partir do máximo – onde o efeito das doses passa a ser mais tóxico do que mutagênico, ajusta-se o decaimento exponencial. Os passos são:

1. Determinar a dose na qual foi observado o maior número de revertentes. Por exemplo, $\tilde{x}$.
2. Ajustar um sub-MLG para obtenção de $\beta_0^{(0)}$ e $\beta_1^{(0)}$, utilizando todos os dados associados às doses menores que $\tilde{x}$. Para o modelo de Stead, utiliza-se perfil de verossi-

milhança para estimar $\beta_2^{(0)}$.

3. Ajustar um sub-MNLG de decaimento exponencial: $e^{-\gamma x}$ para todos os dados associados às doses maiores que $\tilde{x}$ e obter o valor inicial $\gamma^{(0)}$.
4. Ajustar o MNLG completo, utilizando os valores iniciais obtidos nos passos anteriores.

Note que é necessário obter um valor inicial para o submodelo não-linear generalizado de decaimento exponencial, que pode ser feito por uma regressão linear simples com a variável transformada. Há, no entanto, que se considerar a possibilidade de que em muitos experimentos não estão presentes todas as doses que causam o efeito descrito pelo preditor. Se os níveis de x resultaram apenas em efeito tóxico, é difícil obter as estimativas (iniciais ou finais) para os parâmetros associados à mutagenicidade. Por outro lado, se a toxicidade estiver presente como descrito pelo preditor, mas os níveis de x contemplarem apenas a porção da curva em que a mutagenicidade é o efeito predominante, um MLG polinomial não deve apresentar bom ajuste e, dessa forma, um valor inicial para o parâmetro do decaimento exponencial, γ , precisa ser obtido. É natural pensar que a informação sobre tal quantidade esteja presente nos resíduos do MLG polinomial associado, entretanto questões de escala precisam ser consideradas. Nesses casos, duas alternativas à obtenção de $\gamma^{(0)}$ pelos resíduos são: simplesmente propor $\gamma^{(0)} = 0$ ou insistir no ajuste de um sub-MNLG para o decaimento exponencial, que geralmente resulta em estimativas muito próximas de zero. A função `iv.ames()`, na versão atual do pacote, consegue apenas lidar com a segunda situação.

Um segundo estudo de simulação foi conduzido, utilizando os mesmos cenários e valores dos parâmetros da seção 4.1.1, para avaliar a frequência de convergência do algoritmo de estimação quando os valores iniciais são obtidos por sistema de equações ou por ajuste de submodelos, com adição e sem adição de perturbação no vetor de estimativas iniciais.

4.1.3 Ajuste dos modelos tradicionais

Aos dados das Tabelas 5 e 6 foi aplicado o teste de escore Dean (DEAN, 1992) para superdispersão, conforme descrito na seção 2.3.4, pela função, escrita em R, `dtest.ames()` (ver seção 4.5.3).

Posteriormente, foram ajustados, separadamente para cada inseticida e cada SAM S9, os modelos compostos pela combinação entre preditor (Bernstein, Breslow, Krewski, Margolin, Myers ou Stead) e família (Poisson por MV, quase-Poisson por MQV, BN-2 por MV ou quase-potência por MQVE) e, em todos os casos, foi utilizada a função de ligação logarítmica. Essas famílias foram priorizadas de acordo com a forma e complexidade da função de variância, e o método de estimação seguiu aquele apresentado na Tabela 3.

Foram reportados, para os modelos ajustados por MV, o AIC_c , a estatística generalizada X^2 de Pearson e a *deviance* residual. Para os modelos ajustados por MQV ou MQVE, foram reportados o AIC_c e a quase-*deviance* residual escalonada.

Para cada modelo, foram obtidos os gráficos de resíduos: componentes de *deviance* estudentizados (r_i^{d*}) *versus* variáveis explanatórias (doses); componentes de *deviance* estudentizados *versus* preditores ajustados ($\hat{\eta}$) e um gráfico de componentes de *deviance* estudentizados ordenados *versus* quantis da distribuição normal. Neste gráfico, foram incluídos envelopes simulados de acordo com a família ajustada, sendo que para a quase-Poisson os envelopes foram gerados a partir de uma distribuição Poisson generalizada (ZAMANI; ISMAIL, 2012) e, para a quase-potência, a partir de uma distribuição Tweedie discretizada (DUNN, 2014; JØRGENSEN, 1997, p. 127).

Também foram obtidos, para diagnósticos, gráficos de índices de *leverage* e gráficos de influência e curvatura locais, utilizando diversos subconjuntos do vetor de parâmetros de cada preditor.

Os ajustes foram realizados em R (R CORE TEAM, 2015), na função `fit.ames()`, desenvolvida neste trabalho e apresentada na seção 4.5.1. O método de obtenção dos valores iniciais para o algoritmo de estimação foram obtidos seguindo os sistemas de equações, na seção 4.1.2, disponíveis na função `iv.ames()`, mais detalhadamente descrita na seção 4.5.2. Os gráficos e as medidas de diagnósticos foram obtidos respectivamente pelas funções `plot.fitames()` e `diagnostics.fitames()` (ver seções 4.5.4 e 4.5.5).

Uma vez que os modelos tradicionais podem, eventualmente, não se ajustar bem aos dados, foram propostos os modelos da família Skellam (seção 4.2), que são MLGVs, e modelos não-lineares vetoriais da família binomial negativa (seção 4.3).

4.2 Modelos da família Skellam

O modelo da família Skellam, ou regressão de Skellam, é um caso particular de MLGV (seção 2.4), para o qual se supõe que a variável resposta tenha distribuição de Skellam. Trata-se de uma generalização daquela obtida por Irwin (1937) para a diferença de duas variáveis aleatórias com a mesma distribuição Poisson, e foi primeiramente proposta por Skellam (1946), como segue: seja a diferença das variáveis aleatórias $R_1 - R_2$, em que $R_i \sim \text{Poisson}(\mu_i)$. Então, se $Y = R_1 - R_2$, isto implica $Y \sim \text{Skellam}(\mu_1, \mu_2)$, com distribuição de probabilidades dada por:

$$p(Y = y; \mu_1, \mu_2) = e^{-(\mu_1 + \mu_2)} \left(\frac{\mu_1}{\mu_2} \right)^{y/2} I_{|y|}(2\sqrt{\mu_1 \mu_2}); y \in \mathbb{Z}, \quad (49)$$

em que $I_\nu(\cdot)$ é a função modificada de Bessel de primeira espécie. Algumas propriedades da distribuição de Skellam, algumas vezes referida como sendo a “distribuição da diferença de Poisson”, podem ser encontradas no trabalho de Alzaid e Omair (2010).

Do ponto de vista prático, a construção do modelo da família Skellam requer a

especificação dos primeiro e segundo momentos da distribuição representada na equação (49), que são dados, respectivamente, por $E(Y) = \mu_1 - \mu_2$ e $\text{Var}(Y) = \mu_1 + \mu_2$. Isto significa que dois preditores, um para μ_1 e outro para μ_2 , precisam ser especificados.

Esses modelos se apresentam promissores para os dados do TA: para cada concentração avaliada, ocorre um balanço entre os efeitos de mutagenicidade e toxicidade, isto é, o número final de revertentes observados por placa pode ser entendido como a diferença entre o número de UFCs que sofreram mutação, cujo valor esperado é μ_1 , e o número de UFCs que morreram, com valor esperado μ_2 .

Atualmente, no *software* livre R (R CORE TEAM, 2015), há um algoritmo disponível, no pacote “VGAM” (YEE, 2014), para obter as estimativas de um modelo da família Skellam. Esse algoritmo apresenta forte instabilidade numérica, sobretudo para valores grandes de $E(Y)$ (YEE, 2010). Para contornar esse problema, foram propostas neste trabalho as funções `dskellam2()` e `besselIs2()`, que são chamadas internamente na função `fit.ames()` (ver seção 4.5).

No algoritmo da equação (30), a matriz $\mathbf{X}_i$ e o vetor $\mathbf{z}_i$ são fáceis de obter. Entretanto, $\mathbf{W}_i$ tem estrutura complicada para a família Skellam pelo método escore de Fisher. A proposta de Yee (2010) foi de aproximar a informação esperada pela matriz empírica de informação esperada. O método consiste em tomar a média do produto matricial dos escores para os preditores, por meio de simulações de Monte Carlo, como segue: seja $\mathbf{d}_i = \partial L_i / \partial \boldsymbol{\eta}_i$ o vetor escore para os preditores $\boldsymbol{\eta}_1 = \log \mu_1$ e $\boldsymbol{\eta}_2 = \log \mu_2$ na i -ésima observação. Então,

$$\tilde{\mathbf{W}}_i^{(a)} = \frac{1}{T} \sum_{t=1}^T [\mathbf{d}_i^T \mathbf{d}_i]_t \Big|_{\boldsymbol{\eta}_i = \boldsymbol{\eta}_i^{(a)}} \approx \mathbf{W}_i^{(a)},$$

$\tilde{\mathbf{W}}_i^{(a)}$ é uma boa aproximação de $\mathbf{W}_i^{(a)}$ na a -ésima iteração do algoritmo, para um número suficientemente grande T de simulações de Monte Carlo. O processo é repetido para cada uma das observações, $i = 1, 2, \dots, n$.

Neste trabalho, foram consideradas funções polinomiais nas doses, para os preditores $\boldsymbol{\eta}_1$ e $\boldsymbol{\eta}_2$, com e sem interação entre dose e SAM S9, sem intercepto em todos os casos. A ausência do intercepto é justificada pelo já conhecido número de revertentes espontâneos (ver Tabela 1) e, assim, foram adotados modelos com variável *offset*, dada pela mediana do número de revertentes no controle negativo.

4.3 Modelos não-lineares vetoriais da família binomial negativa

Dentro do contexto dos MLGVs (seção 2.4), há uma possibilidade de generalização imediata e ainda pouco estudada, que é eliminar a restrição de linearidade dos preditores. Seguindo as mesmas condições de regularidade dos MLGVs, um modelo não-linear generalizado vetorial é compatível com a equação (29), alterando a definição dos pre-

ditores. Neste caso, cada $\eta_j(\cdot)$ é o preditor potencialmente não-linear do parâmetro λ_j , tal que $j = 1, 2, \dots, m$, satisfazendo $\eta_j = f_j(\boldsymbol{\theta}_j, \mathbf{x}_{ij})$; $f_j(\cdot)$ é o j -ésimo componente determinístico do modelo – uma função tipicamente não-linear, contínua e diferenciável, que relaciona o vetor de covariáveis $\mathbf{x}_{ij}$ de variáveis explanatórias ao vetor de parâmetros desconhecidos $\boldsymbol{\theta}_j$ de comprimento p_j , para a i -ésima observação.

A estimação nesses modelos pode ser feita também por MQRI, bastando iterar a matriz local do experimento na equação (30), obtendo-se o algoritmo modificado

$$\boldsymbol{\theta}^{(a+1)} = \left(\sum_{i=1}^n \tilde{\mathbf{X}}_i^{\top(a)} \mathbf{W}_i^{(a)} \tilde{\mathbf{X}}_i^{(a)} \right)^{-1} \left(\sum_{i=1}^n \tilde{\mathbf{X}}_i^{\top(a)} \mathbf{W}_i^{(a)} \mathbf{z}_i^{*(a)} \right),$$

em que $\tilde{\mathbf{X}}_i^{(a)}$ é a matriz local do experimento na a -ésima iteração do algoritmo, que é bloco-diagonal $m \times \sum_{j=1}^m p_j$, dada por

$$\tilde{\mathbf{X}}_i^{(a)} = \text{diag} \left\{ \left. \frac{\partial \eta_{i1}}{\partial \boldsymbol{\theta}_1} \right|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(a)}}, \left. \frac{\partial \eta_{i2}}{\partial \boldsymbol{\theta}_2} \right|_{\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^{(a)}}, \dots, \left. \frac{\partial \eta_{im}}{\partial \boldsymbol{\theta}_m} \right|_{\boldsymbol{\theta}_m = \boldsymbol{\theta}_m^{(a)}} \right\};$$

$\mathbf{z}_i^{*(a)} = \tilde{\mathbf{X}}_i^{(a)} \boldsymbol{\theta}^{(a)} + \mathbf{W}_i^{-1(a)} \mathbf{d}_i^{(a)}$ são os preditores modificados; $\mathbf{W}_i^{(a)}$ e $\mathbf{d}_i^{(a)}$ são equivalentes aos da equação (30). Este procedimento será primeiramente discutido em Butturi-Gomes e Zocchi³ (em fase de elaboração), que contará com medidas de influência e curvatura locais desenvolvidas para MNLGVs.

Para o caso particular de um componente aleatório BN-2 foram consideradas as funções de ligação $g_1(\boldsymbol{\mu}) = \log \boldsymbol{\mu} = \boldsymbol{\eta}_1$ e $g_2(\mathbf{k}) = \log \mathbf{k} = \boldsymbol{\eta}_2$. Neste cenário, o vetor $\mathbf{d}_i$ tem comprimento 2 dado por:

$$\mathbf{d}_i = \frac{\partial L(\mu_i, k_i; y_i)}{\partial \boldsymbol{\eta}} = \begin{bmatrix} \frac{\partial L(\mu_i, k_i; y_i)}{\partial \eta_{i1}} \\ \frac{\partial L(\mu_i, k_i; y_i)}{\partial \eta_{i2}} \end{bmatrix} = \begin{bmatrix} \frac{d\mu_i}{d\eta_{i1}} \frac{\partial L(\mu_i, k_i; y_i)}{\partial \mu_i} \\ \frac{dk_i}{d\eta_{i2}} \frac{\partial L(\mu_i, k_i; y_i)}{\partial k_i} \end{bmatrix},$$

em que $d\mu_i/d\eta_{i1} = \exp\{\eta_{i1}\}$; $dk_i/d\eta_{i2} = \exp\{\eta_{i2}\}$; $\partial L_i/\partial \mu_i = k_i(y_i - \mu_i)/\mu_i(\mu_i + k_i)$; e $\partial L_i/\partial k_i = 1 + \psi(y_i + k_i) - \psi(k_i) + \log k_i - \log(\mu_i + k_i) - (y_i + k_i)/(\mu_i + k_i)$. Para simplicidade de notação $L_i \equiv L(\mu_i, k_i; y_i)$, e $\psi(\cdot) = \Gamma'(\cdot)/\Gamma(\cdot)$ é a função digama. A matriz $\mathbf{W}_i$ é 2×2 e dada por:

$$-\mathbf{W}_i = \text{E} \begin{bmatrix} \frac{\partial^2 L_i}{\partial \eta_{i1}^2} & \frac{\partial^2 L_i}{\partial \eta_{i1} \partial \eta_{i2}} \\ \frac{\partial^2 L_i}{\partial \eta_{i2} \partial \eta_{i1}} & \frac{\partial^2 L_i}{\partial \eta_{i2}^2} \end{bmatrix} = \begin{bmatrix} \left(\frac{d\mu_i}{d\eta_{i1}} \right)^2 \text{E} \left(\frac{\partial^2 L_i}{\partial \mu_i^2} \right) & 0 \\ 0 & \left(\frac{dk_i}{d\eta_{i2}} \right)^2 \text{E} \left(\frac{\partial^2 L_i}{\partial k_i^2} \right) \end{bmatrix},$$

³BUTTURI-GOMES, D.; ZOCCHI, S.S. **A note on vector generalized linear models: a modified algorithm to allow for nonlinear predictors.**

em que, seguindo os resultados de Lawless (1987, p. 211),

$$E(\partial^2 L_i / \partial k_i \partial \mu_i) = E(\partial^2 L_i / \partial \mu_i \partial k_i) = 0;$$

$$E(\partial^2 L_i / \partial \mu_i^2) = -[k_i \mu_i^2 + k_i^2 \mu_i] [\mu_i^2 (\mu_i + k_i)^2]^{-1};$$

e

$$E(\partial^2 L_i / \partial k_i^2) = -\sum_{t=0}^{+\infty} [(k_i + t)^{-2} P(Y_i \geq t)] - k_i^{-1} \mu_i / (\mu_i + k_i).$$

Por fim, resta apenas especificar a matriz local do modelo $\tilde{\mathbf{X}}_i$, que depende da forma, linear ou não-linear, dos preditores $\boldsymbol{\eta}_1$ e $\boldsymbol{\eta}_2$. Neste trabalho, foram consideradas, para $\boldsymbol{\eta}_1$, as mesmas formas dos preditores tradicionais (equações (5) a (10)). Para $\boldsymbol{\eta}_2$, foram propostas formas lineares polinomiais nas concentrações do inseticida, com e sem interação com o SAM S9. Note que, se $\boldsymbol{\eta}_2 = \theta_{02}$, isto é, um preditor de apenas intercepto, o modelo final é equivalente a um MNLG, com componente aleatório BN-2, para o qual as estimativas do parâmetro k são realizadas por máxima verossimilhança, utilizando equações escore.

4.4 Estratégias para seleção de modelos, inferência e medidas de interesse

Neste trabalho foram estabelecidas estratégias gerais para a seleção, devido ao grande número de modelos sob investigação. Os dois critérios mínimos para que um modelo seja considerado plausível são que haja convergência do algoritmo e que, se houver, as estimativas obtidas estejam dentro da faixa de interpretação biológica. A combinação dos diferentes preditores tradicionais com diferentes componentes aleatórios cria um cenário de modelos não encaixados e, então, a estratégia para a seleção do melhor modelo tradicional se deu pela conferência dos seguintes passos:

- convergência do algoritmo de estimação;
- obtenção de estimativas dentro do espaço paramétrico, respeitando os processos biológicos;
- resultados do teste escore de Dean para superdispersão, realizados unilateralmente a 95% de significância, independentemente para cada preditor e para cada forma da variância total na hipótese alternativa ($V = \mu(1 + \zeta)$ e $V = \mu_i + \zeta \mu_i^2$);
- $\Delta^{(AIC)} \leq FEE$; $FEE=2$, seguindo Burnham e Andersen (2004), como apresentado na Tabela 4, seção 2.5.3.
- teste de qui-quadrado a 95% de significância para bondade de ajuste, se disponível;

- avaliação dos gráficos de resíduos e diagnósticos;
- propriedades das distribuições amostrais dos estimadores dos parâmetros;
- verossimilhança completamente especificada.

Assim, a cada passo, mais e mais modelos foram excluídos no processo de seleção. Uma vez que, dentre todos os modelos tradicionais, os melhores para cada conjunto de dados tivessem sido identificados, foram ajustados os modelos não-lineares vetoriais da família binomial negativa tipo-2 utilizando os preditores tradicionais daqueles pré-selecionados. Desta forma, foi possível avaliar as interações dos parâmetros de mutagenicidade e toxicidade com o SAM S9 e verificar se havia dependência sistemática do parâmetro de superdispersão. Como estes modelos são encaixados, foram realizados testes de razão de verossimilhanças para seleção. Finalmente, foram incluídos os modelos da família Skellam e os passos supracitados foram repetidos, restando apenas um modelo para cada conjunto de dados.

Para cada modelo efetivamente selecionado, foram reportados: as estimativas dos parâmetros, os desvios-padrões assintóticos (DP) e o respectivo intervalo de confiança (IC) assintótico a 95% (IC-1). Também foram obtidos os ICs *bootstrap* baseados em percentis (IC-2), seguindo as recomendações de Efron e Tibshirani (1993, p. 170).

Por fim, baseando-se nesses modelos, foram obtidas as medidas de interesse: a dose de mutagenicidade máxima (DMM), que é simplesmente a dose na qual se obtém o maior valor ajustado para a variável resposta; a fração de mutagenicidade máxima (FMM), dada pela razão entre a DMM e a dose recomendada para uso em campo (DC) do inseticida; o índice de mutagenicidade máxima (IMM), que é a razão entre o maior valor ajustado para a variável resposta e a estimativa do número de revertentes espontâneos, geralmente descrito pelo intercepto do modelo; e o nível de risco ambiental (NRA), obtido pela razão entre o IMM e a FMM. Para cada uma dessas medidas de interesse, foram calculados os ICs *bootstrap* baseados em percentis.

4.5 Implementações computacionais: o pacote “ames”

Com o objetivo de facilitar o ajuste de modelos e a análise de resíduos e diagnósticos, foi desenvolvido um pacote, dedicado a dados provenientes do teste *Salmonella*/microsoma, em linguagem R (R CORE TEAM, 2015). O pacote, intitulado “ames”, é de autoria de Butturi-Gomes e Zocchi⁴ e, em breve, estará disponível para *download* na CRAN (em fase de elaboração). Uma versão do pacote pode ser obtida ao executar `install_github('dbgstat/ames')`, que requer o pacote “devtools” (WICKHAM; CHANG, 2015), diretamente na linha de comando do R.

⁴BUTTURI-GOMES, D.; ZOCCHI, S.S. **ames: statistical solutions for Ames test data**. R package version 0.5-0.

O pacote “ames” possui dependências de outras funções, disponíveis em: “rootSolve” (SOETAERT, 2009; SOETAERT; HERMAN, 2009), “Bessel” (MAECHLER, 2015), “Matrix” (BATES; MAECHLER, 2015), “expm” (GOULET et al., 2014) e “MASS” (VENABLES; RIPLEY, 2002). Os argumentos de entrada e a estrutura de saída das principais funções desenvolvidas estão descritas nas seções 4.5.2 a 4.5.8. Na seção 4.5.9 são apresentadas algumas funções secundárias.

4.5.1 Função `fit.ames()`

Trata-se da função principal do pacote “ames”, que produz os ajustes de todos os modelos investigados no presente trabalho, e também fornece a classe de objetos no R para a análise de resíduos e diagnósticos, de acordo com o modelo ajustado. Alguns métodos estão disponíveis, como `plot(fit.ames())` e `print(fit.ames())`.

`fit.ames(data, count, dose, lh, predictor, link, form1, form2, theta.start, dispersion.start, dispersion.method, controls, quiet)`

- **data**: Objetos do tipo matriz (*matrix*) ou conjunto de dados (*data.frame*). O padrão é contagens na primeira coluna e dose (ou concentração) na segunda coluna. Se estiver ausente, a função irá buscar os argumentos **count** e **dose**.
- **count**: Inteiro (*integer*). Um vetor de comprimento igual ao número de observações, contendo o número de revertentes observados por placa. Se **data** e **count** forem fornecidos na mesma execução da função, **data** tem prioridade.
- **dose**: Numérico (*numeric*). Um vetor de comprimento igual ao número de observações, contendo as doses (ou concentrações) de exposição de cada placa. Se **data** e **dose** forem fornecidos na mesma execução da função, **data** tem prioridade.
- **lh**: Sequência de caracteres (*character*), descrevendo o componente aleatório. Sequências válidas para MNLGs são "poisson", "nb2", "quasipoisson", "quasipower". Estas famílias se diferenciam ou pela especificação completa da verossimilhança ou pela escolha da função de variância: "poisson", assume $\phi = 1$, com função de verossimilhança Poisson completa e $V(\mu) = \mu$; "nb2", assume $\phi = 1$ se **dispersion.method** for "ml", com função de verossimilhança binomial negativa tipo-2 completa e $V(\mu) = \mu(1 + \mu/k)$, mas, se **dispersion.method** for "mpl", assume $\phi = \mu(1 + \mu/k)$ com $V(\mu) = 1$ e a estimação é por pseudo-verossimilhança; "quasipoisson" usa o método de quase-verossimilhança, com $V(\mu) = \mu$ e ϕ é estimado por momentos (estatística generalizada de Pearson); "quasipower" assume ϕ desconhecido com função de variância $V(\mu) = \mu^\tau$, de forma que a estimação é feita por máxima quase-verossimilhança estendida e, se **dispersion.method** for "profile", traça o perfil para estimar τ . Para M(N)LGVs, as sequências válidas são: "skellam" assume componente aleatório Skellam e preditores lineares para μ_1 e μ_2 são especificados via argumentos **form1** e **form2**, e "nlvnb2", que assume componente aleatório binomial negativo tipo-2, preditor para μ geralmente não-linear especificado pelo argumento **predictor** e preditor linear para k , de acordo com a fórmula dada em ou

`form1` ou `form2`.

- **predictor**: Sequência de caracteres, descrevendo a forma do preditor, geralmente não-linear, para a média do número de revertentes. Sequências válidas são "krewski", "myers", "bernstein", "breslow", "stead" e "margolin". Este argumento é ignorado no caso da família "skellam".

link: Função (*function*) ou lista de funções, contendo a(s) função(ões) de ligação. Na versão atual do pacote, apenas a ligação `log` está implementada, que é também o padrão.

- **form1**, **form2**: Fórmulas (*formula*). Para a família "skellam", **form1** e **form2** devem ser fórmulas que descrevem a média do comportamento mutagênico e a média do comportamento tóxico, respectivamente. Para a família "nlvnb2", que resulta em um MNLGV, apenas **form1** ou **form2** deve ser fornecido, descrevendo o comportamento do parâmetro k (**size**) da distribuição binomial negativa tipo-2. Ambos os argumentos são ignorados para as demais famílias.

- **theta.start**: Numérico, de comprimento idêntico ao número de parâmetros no(s) preditor(es). São as estimativas iniciais para o(s) parâmetro(s) do(s) preditor(es). Para MNLGs, o padrão é "smart", que chama internamente a função `iv.ames()` em sua forma padrão (ver seção 4.5.2), com ou sem perturbação, de acordo com **controls**.

- **dispersion.start**: Numérico, de comprimento 1 ou 2, dependendo de **lh**. São as estimativas iniciais do(s) parâmetro(s) de dispersão. O padrão, bastante arbitrário, é 1. Para M(N)LGVs, este argumento é ignorado.

- **dispersion.method**: Sequência de caracteres. As sequências válidas são: "ml" (máxima verossimilhança), "mpl" (máxima pseudo-verossimilhança) e "profile" (perfil de verossimilhança, quase-verossimilhança ou quase-verossimilhança estendida). Os padrões são diferentes, dependendo do argumento **lh**: para "poisson" e "nb2", o padrão é "ml"; e para "quasipower" o padrão é "profile". Este argumento é ignorado para M(N)LGVs e para a família "quasipoisson".

- **controls**: Objetos do tipo lista (*list*). Controla aspectos de convergência do algoritmo, das quantidades arbitrárias de alguns preditores, número inicial de UFCs nas placas, precisão dos perfis de verossimilhança, entre outros. O padrão é chamar internamente a função `control.fitames()` (ver seção 4.5.9).

- **quiet**: Lógico. Se TRUE, todo o comando é executado, mas nenhum valor é impresso na tela. O padrão é FALSE.

A função `fit.ames()` retorna um objeto S3 de classe "fitames", contendo:

\$model: Conjunto de dados com apenas uma linha, apresentando informações sobre o modelo ajustado.

\$coef: Conjunto de dados com as estimativas e respectivos desvios padrões assintóticos, dado o modelo ajustado.

\$convergence: Conjunto de dados contendo informações na convergência do algoritmo, número de iterações, graus de liberdade do resíduo, *deviance* residual e o logaritmo da

função de verossimilhança.

\$predictors: Numérico (MNLG), Matriz (MNLGV). Valores ajustados do modelo na escala da função de ligação.

\$fitted.values: Numérico (MNLG), Matriz (MNLGV). Valores ajustados do modelo na escala da variável resposta.

\$data: Conjunto de dados originais usados para ajuste do modelo.

\$hat: Matriz. Trata-se da matriz chapéu, ou matriz de projeção.

\$residuals: Conjunto de dados. Diversos tipos de resíduos, dependendo da classe do modelo ajustado. Em geral, são *working*, resíduos de Pearson, resíduos de Pearson estudantizados, componentes de *deviance* (apenas MNLG), componentes de *deviance* estudantizados (apenas MNLG).

\$fisher: Matriz de informação esperada na convergência do algoritmo.

\$matrices: Lista de matrizes na convergência do algoritmo, como matriz local do experimento, matriz de pesos e variâncias assintóticas.

\$fit.method: Sequência de caracteres, identificando o algoritmo de estimação pela classe do modelo ajustado.

\$controls: Lista idêntica àquela utilizada como argumento `control` na execução da função.

4.5.2 Função `iv.ames()`

Computa as estimativas iniciais para os parâmetros dos preditores tradicionais, conforme descrito na seção 4.1.2 e retorna um vetor de comprimento igual ao número de parâmetros do preditor escolhido.

```
iv.ames(data, count, dose, predictor, link, controls, interval,
submod, ...)
```

- `data, count, dose, predictor, link, controls:` Argumentos idênticos aos da função `fit.ames()` (seção 4.5.1)

- `interval:` Vetor numérico de comprimento igual a 2. Intervalo de busca para as estimativas iniciais. O padrão é `c(-5,5)`.

- `submod:` Lógico. Se `TRUE` obtém os valores iniciais por ajuste de submodelos. Se `FALSE` (padrão), utiliza sistema de equações.

- `...`: Argumentos extras a serem passados às funções `uniroot` e `uniroot.all`, do pacote “rootSolve” (SOETAERT, 2009).

4.5.3 Função `dtest.ames()`

Executa o teste score para superdispersão de Dean (1992) (ver seção 2.3.4), supondo ajuste de qualquer modelo tradicional para os dados do TA e componente aleatório Poisson. Retorna uma lista e, por padrão, imprime na tela o valor de `p`.

`dtest.ames(alternative.var, alternative, data, count, dose, predictor, link, theta.start, controls)`

- `alternative.var`: Sequência de caracteres, especificando qual a forma da função de variância na hipótese alternativa. Sequências válidas são "gp1" para a alternativa $\phi\mu$ ou "nb2" para a alternativa $\mu(1 + \mu/k)$.
- `alternative`: Sequência de caracteres, indicando o tipo da hipótese alternativa. Sequências válidas são "two.sided", "less" ou "greater". É recomendado o uso da hipótese composta "greater" e, em caso, contrário emite um aviso (*warning*).
- `data, count, dose, predictor, link, theta.start, controls`: Argumentos idênticos aos da função `fit.ames()` (seção 4.5.1). Aceita no lugar de `data` um objeto de classe "fitames" pré-ajustado com componente aleatório Poisson e, neste caso, ignora os demais argumentos para ajuste.

`$Test`: Conjunto de dados com uma linha, descrevendo a forma da variância alternativa, a estatística do teste e o valor de p. A estatística do teste é obtida pela soma dos componentes e dividido pela variância e, sob suposição, tem distribuição normal padrão.

`$Tis`: Vetor numérico, contendo os componentes da estatística do teste.

`$Tis.Var`: Numérico, igual à variância assintótica do teste.

4.5.4 Função `plot.fitames()`

Produz uma série de gráficos para análise de resíduos e diagnósticos de um modelo previamente ajustado por `fit.ames()`. Por padrão, retorna um objeto vazio (*NULL*).

`plot.fitames(obj, smooth, deviance.plots, std.deviance.plots, pearson.plots, std.pearson.plots, qqplot, lag.plot, lag, diag.measures, envelope, nsim, probs, all, quiet)`

- `obj`: Objeto de classe "fitames", geralmente produzido por `fit.ames()`.
- `smooth`: Lógico. Se TRUE produz um gráfico de dispersão dos valores observados *versus* as doses e inclui uma curva suave dos valores ajustados. Se FALSE, produz um gráfico de dispersão dos valores observados *versus* as doses e inclui os pontos dos valores ajustados. O padrão é TRUE.
- `deviance.plots`: Lógico. Se TRUE produz um gráfico de componentes de *deviance* *versus* valores ajustados, outro gráfico *versus* preditores ajustados, e, por último, um gráfico desses componentes *versus* a variável explicativa dose. Se FALSE (padrão) não produz gráfico algum. Esse argumento é ignorado no caso de M(N)LGVs.
- `std.deviance.plots`: Lógico. Se TRUE produz os mesmos tipos de gráficos que `deviance.plots`, mas utiliza os componentes de *deviance* estudentizados. Se FALSE (padrão) não produz gráfico algum. Esse argumento é ignorado no caso de M(N)LGVs.
- `pearson.plots`: Lógico. Se TRUE produz os mesmos tipos de gráficos que `deviance.plots`, mas utiliza os resíduos de Pearson. Se FALSE (padrão) não produz gráfico algum. Se o modelo ajustado for um M(N)LGV, todos os tipos de gráficos serão

produzidos para cada preditor.

- `std.pearson.plots`: Lógico. Se TRUE produz os mesmos tipos de gráficos que `deviance.plots`, mas utiliza os resíduos de Pearson estudentizados. Se FALSE (padrão) não produz gráfico algum. Se o modelo ajustado for um M(N)LGV, todos os tipos de gráficos serão produzidos para cada preditor.
- `lag.plot`: Lógico. Se TRUE produz um gráfico de defasagem residual (*lag plot*). Um gráfico Será produzido para cada tipo de resíduo, de acordo com os argumentos lógicos `deviance.plots`, `std.deviance.plots`, `pearson.plots` e `std.pearson.plots`. Se todos esses forem FALSE, são utilizados os resíduos de Pearson. O padrão é FALSE.
- `lag`: Inteiro. Determina o tamanho da defasagem, se `lag.plot` for TRUE. O padrão é 1.
- `diag.measures`: Lógico. Se TRUE, a função `diagnostics.fitames()` é chamada internamente. O padrão é FALSE.
- `qqplot`: Lógico. Se TRUE Produz gráfico de resíduos ordenados *versus* os quantis da distribuição normal. Será produzido um gráfico para cada tipo de resíduo, de acordo com os argumentos lógicos `deviance.plots`, `std.deviance.plots`, `pearson.plots` e `std.pearson.plots`. Se todos esses forem FALSE, são utilizados os resíduos de Pearson. O padrão é FALSE.
- `envelope`: Lógico. Se TRUE produz um gráfico de resíduos ordenados *versus* os quantis da distribuição normal com envelope simulado, seguindo as mesmas condições do argumento `qqplot`. O padrão é FALSE.
- `nsim`: Inteiro, denotando o número de simulações para o envelope simulado. Este argumento é ignorado se `envelope` for FALSE. O padrão é 99.
- `probs`: Vetor numérico com dois elementos, denotando os limites inferior e superior do envelope simulado. Este argumento é ignorado se `envelope` for FALSE. O padrão é `c(0.005,0.995)`.
- `all`: Lógico. Se TRUE impõe que todos os argumentos lógicos da função sejam TRUE. O padrão é FALSE.
- `quiet`: Lógico. Se FALSE imprime na tela uma lista contendo as informações do envelope simulado e das medidas de diagnóstico. O padrão é TRUE.

4.5.5 Função `diagnostics.fitames()`

Calcula as medidas de influência e curvatura locais e também constrói gráficos de índices de *leverage* para um modelo previamente ajustado por `fit.ames()`.

`diagnostics.fitames(obj, lmaxes.plot, curvs.plot, nuisance.plot, eta.plot)`

- `obj`: Objeto de classe “fitames”, geralmente produzido por `fit.ames()`.
- `lmaxes.plot`: Lógico. Se TRUE (padrão) produz o gráfico de influência local de acordo com os subconjuntos de parâmetros de interesse.
- `curvs.plot`: Lógico. Se TRUE (padrão) produz gráficos de curvatura local para a

média da função de verossimilhança.

- `nuisance.plot`: Lógico. Se TRUE (padrão) produz gráficos de curvatura local para o parâmetro de superdispersão, se disponível.

- `eta.plot`: Numérico. Para M(N)LGVs estabelece para qual(is) preditor(es) os gráficos devem ser construídos. O padrão NULL indica que todos os gráficos devem ser construídos. Este argumento é ignorado no caso de MNLGs.

`$Delta`: Matriz. Contém a matriz de derivadas segundas da verossimilhança perturbada (ver seções 2.3.3 e 2.4.3).

`$L`: Matriz de informação observada.

`$B1`: Lista de matrizes (MNLGV) ou Matriz (MNLG). Contém a matriz de informação observada, restrita aos (conjuntos de) parâmetros de interesse. Se não houver (conjuntos de) parâmetros *nuisance*, será um objeto vazio.

`$Ci`: Matriz de curvaturas locais, cujo número de colunas depende do número de (conjuntos de) parâmetros de interesse.

4.5.6 Função `aic.fitames()`

Extraí o AIC, AIC_c , QAIC ou o $QAIC_c$ de um ou mais modelos ajustados por `fit.ames()`. Computa os valores de AIC, AIC_c , QAIC ou de $QAIC_c$ escalonados (ver seção 2.5.3), se mais de um modelo for fornecido. Neste caso, também constrói um gráfico de índices com os pesos da Akaike.

`aic.fitames(...,C,wis.plot,emql.all,SEE)`

...: Um ou mais objetos de classe “fitames”, geralmente produzidos por `fit.ames()`. Um erro é causado se modelos com verossimilhança completamente especificada e modelos de quase-verossimilhança estiverem juntos na entrada da função.

`C`: Lógico. Se FALSE, o critério é corrigido para viés de primeira ordem. Se TRUE (padrão), o critério é corrigido para viés de segunda ordem.

`wis.plot`: Lógico. Se TRUE, constrói um gráfico com os pesos de Akaike e destaca aqueles cujo $\Delta^{(AIC)}$ é menor que `SEE`. O padrão é FALSE.

`emql.all`: Lógico. Se TRUE, calcula o critério de informação baseado na máxima quase-verossimilhança estendida. Se FALSE (padrão) os cálculos são feitos de acordo com o método de estimação dos objetos de entrada.

`SEE`: Numérico. Determina a força de evidência empírica, segundo a qual um modelo é selecionado ou não de acordo com o conjunto de modelos da entrada. Quanto maior o valor de `SEE`, mais modelos serão selecionados. O padrão é 2.

`$summary`: Conjunto de dados, contendo os modelos ordenados pelos valores de AIC, com uma coluna lógica de seleção de acordo com o $\Delta^{(AIC)}$ calculado e o `SEE` fornecido.

`$aics`: Numérico ou Vetor numérico.

`$delta.aics`: Vetor numérico. Vazio, se apenas um modelo for fornecido.

`$akaike.wis`: Vetor numérico. Vazio, se apenas um modelo for fornecido.

4.5.7 Função `mmd.fitames()`

Obtém a dose de mutagênese máxima (DMM) para cada modelo ajustado ou a DMM multimodelo utilizando os pesos de Akaike.

```
mmd.fitames(..., model.names, akaike.weights, emql.all)
```

- ...: Um ou mais objetos de classe “fitames”, geralmente produzidos por `fit.ames()`.
- `model.names`: Sequência de caracteres (opcional) especificando nomes para os modelos.
- `akaike.weights`: Lógico. Se TRUE, chama internamente a função `aic.fitames`, recupera os pesos de Akaike e usa os valores para retornar o valor da DMM multimodelo. O padrão é FALSE.
- `emql.all`: Lógico. Argumento a ser passado a `aic.fitames`.

4.5.8 Função `rames()`

Simulação pseudo-aleatória de inteiros, representando aqui as contagens do número de revertentes por placa, segundo pressupostos fixos (forma da função do preditor e função de ligação) e aleatórios (distribuição de probabilidades da variável resposta).

```
rames(lh, predictor, link, theta, dispersion, dose, mu.only, controls, quiet)
```

- `lh`: Sequência de caracteres. As sequências válidas são "poisson", que gera números aleatórios (contagens em cada placa) com distribuição Poisson; "nb2", com distribuição binomial negativa tipo-2; e "gp1", com distribuição Poisson generalizada.
- `predictor, link, dose, controls`: Idênticos aos da função `fit.ames()` (ver seção 4.5.1)
- `theta`: Vetor numérico de comprimento correspondente ao número de parâmetros presentes em `predictor`.
- `dispersion`: Numérico. Parâmetro de dispersão para as distribuições binomial negativa tipo-2 (k) e Poisson generalizada (ϕ). Se `lh` for "poisson", este argumento é ignorado.
- `mu.only`: Lógico. Se TRUE a função retorna apenas os valores esperados. O padrão é FALSE.
- `quiet`: Lógico. Se FALSE imprime na tela a saída da função, um objeto de classe “sim.ames”. O padrão é TRUE.

`$data`: Conjunto de dados, com duas colunas. A primeira contém as contagens e, a segunda, as doses.

`$theta`: Vetor numérico. Parâmetros do preditor usados para a simulação.

`$dispersion`: Numérico. Parâmetro de dispersão usado para a simulação.

`$lh`: Sequência de caracteres, indicando a distribuição utilizada para a simulação.

`$predictor`: Sequência de caracteres, indicando qual dos preditores tradicionais foi usado para a média das contagens.

`$link`: Função. Contém a função de ligação entre o parâmetro de média e o preditor. Nesta versão do pacote, apenas a função logarítmica está disponível.

`$controls`: Lista, geralmente gerada por `control.fitames()`.

4.5.9 Funções de apoio

`control.fitames(convergence.tol, max.it, zerocounts, cfu, alpha, delta, allow.jitter, max.jitter, dec, tau.profile, conv.crit, halfstepping, stepsize, diagW.min, detas.method, nsim)`. Controla aspectos da função `fit.ames()` (seção 4.5.1) e retorna uma lista.

- `convergence.tol`: Numérico. É o valor mínimo que o critério de convergência deve atingir antes de que as iterações parem. O padrão é 10^{-6} .
- `max.it`: Inteiro, que estabelece o número máximo de iterações do algoritmo. O padrão é 30.
- `zerocounts`: Numérico, indicando a constante que deve ser somada a contagens nulas. O padrão é 10^{-3} .
- `cfu`: Numérico, indicando o número de UFCs presentes em cada placa no início do experimento, que é particularmente importante no preditor de Margolin. O padrão é 10^8 e igual para todas as placas.
- `alpha`: Numérico. Constante arbitrária α presente no preditor de Krewski. O padrão é 2.
- `delta`: Numérico. Constante arbitrária δ no preditor de Breslow. O padrão é 1, mas existem recomendações de utilizar a menor dose positiva no experimento.
- `allow.jitter`: Lógico. Se TRUE (padrão) permite pequenas perturbações aleatórias nos valores iniciais.
- `max.jitter`: Inteiro. Número máximo de tentativas de convergência via perturbação aleatória nos valores iniciais. O padrão é 20.
- `dec`: Inteiro. Número de casas decimais no método de impressão em tela. O padrão é de 4 casas decimais.
- `tau.profile`: Sequência numérica, indicando os valores de τ para construção do perfil, pelo método MQVE para a família "quasipower". O padrão é de com 1000 valores equiespaçados entre 0 a 4, dados por `seq(0,4,length.out=1000)`.
- `conv.crit`: Sequência de caracteres, especificando o critério de convergência para o algoritmo. As sequências válidas são "likelh"(padrão), "theta" e "dev", que, respectivamente, se referem às diferenças absolutas do logaritmo da função de verossimilhança, à soma dos módulos das diferenças entre os pares das estimativas dos parâmetros, e ao módulo da diferença entre as funções *deviance*.
- `halfstepping`: Lógico. Se TRUE (padrão), permite o uso do fracionamento de passos durante o processo iterativo de estimação em M(N)LGVs. Este argumento é ignorado para MNLGs.
- `stepsize`: Numérico, que define o denominador na fração do tamanho do passo. O padrão é 2. É ignorado para MNLGs e também se `halfstepping` for FALSE.

- `diagW.min`: Numérico. Define o menor valor suficientemente positivo nas diagonais de uma matriz positiva-definida para computar sua inversa. Estabelecer uma maior precisão maior que `.Machine$double.eps^0.75` (padrão) pode causar *underflow*.
- `detas.method`: Numérico/Lógico. Para a família "skellam", define qual método numérico será usado para a aproximação das funções de Bessel. Se for igual a 1, então a função `besselI()` será chamada internamente, disponível na base do R. Se for igual a 2, então função `besselIs2()` será chamada. O método 1 é menos preciso e mais rápido quando comparado ao método 2. O padrão é 1.
- `nsim`: Inteiro. Número de simulações realizadas em cada iteração do algoritmo para computar a matriz empírica de informação esperada. Nesta versão do pacote, é usada apenas para as família "skellam".

`dgenpois(x,mu,phi)` e `rgenpois(n,mu,phi)`. Densidade e geração de números aleatórios da distribuição Poisson generalizada, com valor esperado μ e variância $\phi\mu$.

- `x`: Inteiro ou vetor de inteiros não-negativos para o qual se deseja obter a densidade.
- `n`: Inteiro. Número de realizações da distribuição Poisson generalizada.
- `mu`: Numérico positivo. Parâmetro de média da distribuição Poisson generalizada.
- `phi`: Numérico positivo. Parâmetro de dispersão da distribuição Poisson generalizada.

`dskellam2(x, mu1, mu2, log)`. Densidade da distribuição de Skellam, cuja esperança é $\mu_1 - \mu_2$ e tem variância dada por $\mu_1 + \mu_2$. Esta é uma modificação da função `dskellam()` (YEE, 2014) para aumentar sua precisão.

- `x`: Inteiro ou vetor de inteiros para o qual se deseja obter a densidade.
- `mu1`: Numérico positivo. Parâmetro da distribuição de Skellam.
- `mu2`: Numérico positivo. Parâmetro da distribuição de Skellam.
- `log`: Lógico. Se TRUE, retorna as densidades em escala logarítmica. O padrão é FALSE.

`besselIs2(x, nu, nterm, expon.scaled, log, Ceps)`. Calcula uma aproximação da função modificada de Bessel de primeira espécie $I_\nu(x)$. Esta é uma modificação da função `besselIs()` (MAECHLER, 2015) apenas para permitir que ν seja um vetor.

- `x`: Numérico ou vetor numérico.
- `nu`: Numérico ou vetor numérico.
- `nterm`: Inteiro, que indica o número de termos a serem usados para a aproximação.
- `expon.scaled`: Lógico. Se TRUE, o valor de x é usado na escala $\exp\{-|x|\}$. O padrão é FALSE.
- `log`: Lógico. Se TRUE, retorna a aproximação de $I_\nu(x)$ em escala logarítmica. O padrão é FALSE.
- `Ceps`: Numérico, que denota a tolerância da aproximação e é utilizado para checar se o `nterm` escolhido foi suficiente. O padrão é $8 \cdot 10^{-16}$.

5 RESULTADOS E DISCUSSÃO

5.1 Modelos tradicionais

5.1.1 Estudos de simulação e problema do valor inicial

Os testes de hipótese em cada cenário relativos aos parâmetros β_0 e β_1 do preditor de Bernstein estão apresentados nas Figuras 3 e 4, respectivamente. As distribuições amostrais de ambos os estimadores apresentaram, em geral, boas propriedades assintóticas, isto é, tiveram a normalidade assegurada mesmo com tamanho amostral modesto. Apenas no cenários da PG, nos quais foi observado um padrão ascendente entre a estatística de viés para $\hat{\beta}_1$ e o valor real do parâmetro.

Esse mesmo tipo de padrão também ocorreu, principalmente nos cenários de maior superdispersão da PG ($\phi = 8$), nas estatísticas de viés e de assimetria para $\hat{\beta}_0$ do preditor de Breslow (Figura 5). Deste preditor, o estimador $\hat{\beta}_1$ (Figura 6) apresentou viés constante, independente do nível de superdispersão, da forma da função de variância ou do valor real do parâmetro, mas as estatísticas de excesso de variância, assimetria e curtose se apresentaram, em geral, dentro das bandas de confiança a 95%. As estatísticas para o estimador do parâmetro associado à toxicidade γ (Figura 7) apresentaram padrão ascendente com o valor real do parâmetro nos cenários da PG; a estatística de viés também apresentou esse padrão nos cenários da Poisson e BN-2.

Os estimadores dos parâmetros do preditor de Krewski não apresentaram, em geral, problemas com relação à normalidade de suas distribuições amostrais (Figuras 8 a 10). Apenas nos cenários com maior nível de superdispersão, tanto na PG quanto na BN-2, foram observadas evidências significativas a 95% de significância contra a normalidade, mas nenhum desses casos apresentou valores extremos nas estatísticas para viés, excesso de variância, assimetria ou curtose.

Independente do nível de superdispersão e da forma da função de variância, o estimador $\hat{\beta}_0$ no preditor de Margolin apresentou viés acentuado (Figura 11), descendente com o valor real do parâmetros, nos cenários Poisson e BN-2, e ascendente nos cenários da PG. Os estimadores $\hat{\beta}_1$ e $\hat{\gamma}$ também apresentaram vieses significativos (Figuras 12 e 13, respectivamente), em geral com padrões descendentes em relação ao valores verdadeiros dos parâmetros. Foram identificados alguns casos no cenário de maior superdispersão da PG nos quais a distribuição de $\hat{\beta}_1$ é assimétrica positiva. Para $\hat{\gamma}$, foi observada uma tendência ascendente de assimetria positiva, independente do nível de superdispersão na PG e, opositivamente na BN-2, o padrão observado foi descendente à medida em que aumentavam os valores reais de γ , em direção a assimetrias negativas.

Em geral, as distribuições amostrais dos estimadores parâmetros do preditor de Myers (Figuras 14 a 16) apresentaram bom comportamento, apesar de uma leve assimetria positiva no cenário de maior superdispersão da BN-2, particularmente para $\hat{\gamma}$.

Por fim, os piores desempenhos, isto é, falta de normalidade das distribuições amostrais dos estimadores dos parâmetros se deram nos cenários do preditor de Stead (Figuras 17 a 20). Há diversos casos de valores extremos para as estatísticas avaliadas em todos os estimadores, sobretudo para $\hat{\beta}_2$, que, em todos os cenários da PG, apresentou excesso de variância e padrão ascendente com o valor real do parâmetro, independente do nível de superdispersão. Para os estimadores dos parâmetros de um modelo com este preditor, recomenda-se o uso dos métodos de correção de viés e de assimetria extensivamente discutidos em Cavalcanti et al. (2009), Cordeiro e McCullagh (1991), Cordeiro, Previdelli e Samohyl (2008) e Paula (1992).

Os resultados dos estudos de simulação realizado para avaliar o desempenho dos métodos de obtenção do valor inicial podem ser visualizados na Tabela 8, tratando-se da frequência média do sucesso de convergência do algoritmo de estimação para cada preditor tradicional. Essas médias se referem às frequências de sucesso obtidas para 1000 simulações de cada um dos cenários gerados pela combinação de (vetor de parâmetros $\times$ componente aleatório $\times$ nível de superdispersão), totalizando 700.000 simulações para cada preditor.

Observou-se que o método de sistema de equações teve melhor desempenho, comparando-se ao método do ajuste de submodelos, particularmente na presença de perturbações no vetor de estimativas iniciais. É necessário ressaltar, entretanto, que todos os métodos aplicados para o preditor de Stead foram insatisfatórios, chegando a quase 40% de falha nos diversos cenários.

Tabela 8 - Frequência média (em %) do sucesso de convergência do algoritmo de estimação para cada preditor tradicional utilizando os métodos de obtenção de valores iniciais, por sistema de equações (SQ) ou por ajustes de submodelos (SM), na ausência (A) e na presença (P) de perturbações

Preditor	SQ-A	SM-A	SQ-P	SM-P
Krewski	100	100	100	100
Margolin	99,98	87,73	100	87,74
Myers	100	83,36	100	83,4
Stead	70,19	63,43	71,82	63,47

Figura 3 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_0$ do preditor de Bernstein em cada cenário. $\circ$ Equidispersão; $\square$ $\phi = 2$ para PG ou $k = 100$ para BN-2; $\bullet$ $\phi = 4$ para PG ou $k = 50$ para BN-2; $\triangle$ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 4 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_1$ do preditor de Bernstein em cada cenário. $\circ$ Equidispersão; $\square$ $\phi = 2$ para PG ou $k = 100$ para BN-2; $\bullet$ $\phi = 4$ para PG ou $k = 50$ para BN-2; $\triangle$ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 5 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_0$ do preditor de Breslow em cada cenário. ○ Equidispersão; □ $\phi = 2$ para PG ou $k = 100$ para BN-2; ● $\phi = 4$ para PG ou $k = 50$ para BN-2; △ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 6 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_1$ do preditor de Breslow em cada cenário. $\circ$ Equidispersão; $\square$ $\phi = 2$ para PG ou $k = 100$ para BN-2; $\bullet$ $\phi = 4$ para PG ou $k = 50$ para BN-2; $\triangle$ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 7 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\gamma}$ do preditor de Breslow em cada cenário. ○ Equidispersão; □ $\phi = 2$ para PG ou $k = 100$ para BN-2; ● $\phi = 4$ para PG ou $k = 50$ para BN-2; △ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 8 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_0$ do preditor de Krewski em cada cenário. $\circ$ Equidispersão; $\square$ $\phi = 2$ para PG ou $k = 100$ para BN-2; $\bullet$ $\phi = 4$ para PG ou $k = 50$ para BN-2; $\triangle$ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 9 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_1$ do preditor de Krewski em cada cenário. $\circ$ Equidispersão; $\square$ $\phi = 2$ para PG ou $k = 100$ para BN-2; $\bullet$ $\phi = 4$ para PG ou $k = 50$ para BN-2; $\triangle$ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 10 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\gamma}$ do preditor de Krewski em cada cenário. $\circ$ Equidispersão; $\square$ $\phi = 2$ para PG ou $k = 100$ para BN-2; $\bullet$ $\phi = 4$ para PG ou $k = 50$ para BN-2; $\triangle$ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 11 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_0$ do preditor de Margolin em cada cenário. $\circ$ Equidispersão; $\square$ $\phi = 2$ para PG ou $k = 100$ para BN-2; $\bullet$ $\phi = 4$ para PG ou $k = 50$ para BN-2; $\triangle$ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 12 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_1$ do preditor de Margolin em cada cenário. $\circ$ Equidispersão; $\square$ $\phi = 2$ para PG ou $k = 100$ para BN-2; $\bullet$ $\phi = 4$ para PG ou $k = 50$ para BN-2; $\triangle$ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 13 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\gamma}$ do preditor de Margolin em cada cenário. $\circ$ Equidispersão; $\square$ $\phi = 2$ para PG ou $k = 100$ para BN-2; $\bullet$ $\phi = 4$ para PG ou $k = 50$ para BN-2; $\triangle$ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 14 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_0$ do preditor de Myers em cada cenário. $\circ$ Equidispersão; $\square$ $\phi = 2$ para PG ou $k = 100$ para BN-2; $\bullet$ $\phi = 4$ para PG ou $k = 50$ para BN-2; $\triangle$ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 15 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_1$ do preditor de Myers em cada cenário. $\circ$ Equidispersão; $\square$ $\phi = 2$ para PG ou $k = 100$ para BN-2; $\bullet$ $\phi = 4$ para PG ou $k = 50$ para BN-2; $\triangle$ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 16 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\gamma}$ do preditor de Myers em cada cenário. $\circ$ Equidispersão; $\square$ $\phi = 2$ para PG ou $k = 100$ para BN-2; $\bullet$ $\phi = 4$ para PG ou $k = 50$ para BN-2; $\triangle$ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 17 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_0$ do preditor de Stead em cada cenário. ○ Equidispersão; □ $\phi = 2$ para PG ou $k = 100$ para BN-2; ● $\phi = 4$ para PG ou $k = 50$ para BN-2; △ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 18 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_1$ do preditor de Stead em cada cenário. $\circ$ Equidispersão; $\square$ $\phi = 2$ para PG ou $k = 100$ para BN-2; $\bullet$ $\phi = 4$ para PG ou $k = 50$ para BN-2; $\triangle$ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 19 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\beta}_2$ do preditor de Stead em cada cenário. $\circ$ Equidispersão; $\square$ $\phi = 2$ para PG ou $k = 100$ para BN-2; $\bullet$ $\phi = 4$ para PG ou $k = 50$ para BN-2; $\triangle$ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

Figura 20 - Estatísticas dos testes de hipótese sobre a normalidade assintótica da distribuição amostral de $\hat{\gamma}$ do preditor de Stead em cada cenário. ○ Equidispersão; □ $\phi = 2$ para PG ou $k = 100$ para BN-2; ● $\phi = 4$ para PG ou $k = 50$ para BN-2; △ $\phi = 8$ para PG ou $k = 25$ para BN-2; linhas tracejadas: regiões de aceitação das hipóteses nulas a 95% de significância

5.1.2 Ajuste dos modelos

A lógica na apresentação dos resultados nesta seção (5.1.2) e na seção 5.2 consiste em abordar cada passo do processo de seleção na ordem estabelecida na seção 4.4, primeiro para os dados obtidos nos experimentos com a linhagem TA98 e depois para aqueles obtidos com a linhagem TA100. Dentro dos resultados para cada linhagem, os resultados são apresentados primeiro para aqueles obtidos sob exposição ao inseticida Fipronil e, depois, para o Tiametoxam. Finalmente, dentro dos resultados para inseticida em cada linhagem, são apresentados primeiro para a ausência do sistema de ativação metabólica S9 e depois para a presença de S9.

Houve sucesso de convergência do algoritmo de estimação para a maior parte dos modelos utilizando os preditores tradicionais ajustados aos dados da Tabelas 5 e 6, excetuando-se os dados da linhagem TA98 exposta ao inseticida Tiametoxam, que tiveram as piores frequências de sucesso de convergência.

Para os dados do experimento com a linhagem TA98 exposta ao inseticida Fipronil, na ausência e na presença de S9, apenas o preditor de Breslow foi descartado, uma vez que a convergência se deu fora do espaço de interesse biológico, isto é, as estimativas dos parâmetros associados às ações mutagênica e tóxica não possuíam interpretação, com $\beta_1 < 0$ e $\gamma < 0$, conforme apresentado na seção 2.2. Para os dados do experimento com esse inseticida na presença de S9, todos os modelos ajustados utilizando o preditor de Stead não convergiram.

Ainda com relação à linhagem TA98, apenas os modelos com os preditores de Bernstein e Stead, ajustados aos dados do experimento de exposição ao Tiametoxam, atingiram sucesso na convergência e apresentaram estimativas adequadas, do ponto de vista biológico, para os parâmetros.

Para a linhagem TA100 exposta ao inseticida Fipronil, os modelos com os seguintes preditores falharam ou na convergência ou na produção de estimativas com interpretação biológica: Breslow, Krewski e Stead. Finalmente, para os dados do experimento com essa linhagem e o inseticida Tiametoxam, houve sucesso de convergência em todos os demais modelos, excetuando-se aqueles ajustados com o preditor de Bernstein, cujas estimativas dos parâmetros indicaram a presença exclusiva de efeito tóxico ($\beta_1 < 0$). Ainda que seja possível que uma substância produza apenas esse tipo de efeito, uma simples verificação da dispersão dos dados obtidos no presente trabalho (Figuras 21 e 22) é suficiente para descartar essa ideia.

Os resultados do teste score de Dean para superdispersão estão apresentados nas Tabelas 9 e 10 respectivamente para os modelos ajustados aos dados da linhagem TA98 exposta aos inseticidas Fipronil e Tiametoxam, e nas Tabelas 11 e 12, referentes aos modelos ajustados aos dados da linhagem TA100 com os inseticidas Fipronil e Tiametoxam.

Como esperado, a hipótese de equidispersão é geralmente rejeitada nos dados dos

experimentos com ambas as linhagens de *S. typhimurium*. Considerando, entretanto, os modelos ajustados aos dados da linhagem TA98 exposta ao inseticida Fipronil e na presença de S9, a hipótese de equidispersão não foi rejeitada em nenhum caso, independente da forma da variância total na hipótese alternativa, e o mesmo pode ser afirmado para os modelos ajustados aos dados dessa linhagem exposta ao inseticida Tiametoxam na ausência de S9.

Todos esses testes, entretanto, dizem muito pouco a respeito da adequação de um modelo aos dados e, assim, devem ser utilizados em conjunto com medidas apropriadas de seleção e de bondade de ajuste. Levando em consideração as medidas de $\Delta^{(AIC)}$, os modelos pré-selecionados para os dados do experimento com a linhagem TA98, exposição ao inseticida Fipronil e ausência de S9, foram aqueles com preditor de Krewski, com componentes aleatórios BN-2 ou quase-Poisson; com preditor de Myers, com componentes aleatórios BN-2 ou quase-Poisson; e com preditor de Margolin com componente aleatório quase-Poisson (Tabela 13). Dentre estes cinco modelos, os dois para os quais teste de bondade de ajuste estão disponíveis, isto é, têm verossimilhança completamente especificada, não houve evidência estatística suportando a hipótese alternativa de falta de ajuste.

Note que a medida $\Delta^{(AIC)}$ faz uso dos valores de AIC_c e uma notação mais adequada poderia ser $\Delta^{(AIC_c)}$. No trabalho de Hurvich e Tsai (1995), as medidas de AIC_c são obtidas a partir de componentes da função de quase-verossimilhança estendida em modelos encaixados. Pelo fato de serem aninhados, em um contexto para seleção de covariáveis, o problema do viés na estimação da informação de K-L (ver seção 2.5.3) é eliminado e, conseqüentemente, também nos valores de $\Delta^{(AIC_c)}$. Segundo Anderson⁵, tratando-se de modelos não-encaixados como os do presente trabalho, a medida $\Delta^{(\cdot)}$ deveria fazer uso do QAIC, baseado nas funções de quase-verossimilhança (informação pessoal). Nesta situação, além de não anular o viés na estimação da informação de K-L, os modelos da família “quase” ou ganham vantagem injusta para $\phi \gg 1$ ou sofrem desvantagem excessiva para $\phi < 1$.

Devido a esse contraste de ideias e sem mais literatura específica disponível, foi considerado mais justo o uso de $QAIC_c^+$ baseado na funções de quase-verossimilhança estendidas completas, obtido pelo argumento `emql.all=T` na função `aic.fitames()` (ver seção 4.5.6). Assim, a notação mais clara para as medidas de diferenças de $QAIC_c^+$ é $\Delta^{(QAIC_c^+)}$. No presente trabalho todas as referências à $\Delta^{(AIC)}$ para os modelos tradicionais devem ser entendidas como de fato sendo $\Delta^{(QAIC_c^+)}$, ressaltando que estudos sobre as propriedades de $QAIC_c^+$ ainda estão pendentes.

O próximo critério para a seleção dentre os cinco pré-selecionados, é a análise de resíduos e diagnósticos. Nas Figuras 23 a 27 estão representadas as medidas de interesse, respectivamente para os modelos de Krewski com componente aleatório BN-2, de Krewski

⁵ANDERSON, D.R. **AICc and EMQL**. Mensagem recebida por <davi.butturi@gmail.com> em 31 jul. 2015.

quase-Poisson, de Margolin quase-Poisson, de Myers BN-2 e de Myers quase-Poisson.

Os gráficos de *deviance* estudentizada, r_i^{d*} , *versus* dose do inseticida Fipronil apresentaram o padrão nulo para todos os modelos. Já os gráficos de r_i^{d*} *versus* preditores ajustados, $\hat{\eta}_i$, apresentaram em todos os cinco modelos, resíduos concentrados em valores pequenos dos preditores ajustados e, apesar da amplitude constante, é necessário maior número de pontos de suporte em torno dos maiores valores de $\hat{\eta}_i$ para confirmar o padrão nulo.

Os gráficos de resíduos ordenados com envelope simulado, $r_{(i)}^{D*}$, *versus* quantis da distribuição normal, apresentaram forma em “S”, implicando em uma distribuição de caudas mais pesadas. Nenhum resíduo, contudo, se apresentou fora dos envelopes simulados, ainda que os modelos com componente aleatório BN-2 tenham apresentado um ligeiro melhor condicionamento, levando em consideração a distribuição dos resíduos em torno da reta com coeficiente angular igual a 1.

Com relação às medidas de diagnóstico, não foram observados valores de *leverage*, h_{ii} , excessivos para nenhum dos modelos, e sobre a influência ($|l_{\max}|$) e curvatura locais (C_i) apenas observações do controle negativo mostraram valores altos. Estes valores altos, adotando como limites nominais aqueles utilizados por Svetliza e Paula (2003), não são necessariamente um indicativo de falta de ajuste, uma vez que todas as observações e os valores estimados para o intercepto de cada modelo estão dentro dos valores considerados adequados (ver Tabela 1 e o trabalho de Mortelmans e Zeiger (2000)).

Pela análise de resíduos e diagnósticos, nenhum dos cinco modelos pré-selecionados pode ser prontamente descartado. Comparando-se, então, os comportamentos das distribuições amostrais dos estimadores dos parâmetros de cada modelo, deve-se selecionar aqueles com preditor de Myers, que, conforme discutido na seção 5.1.1, é o que apresenta, dentre os pré-selecionados, as melhores propriedades de normalidade, independentemente da forma do componente aleatório. É claro que há vantagens de se trabalhar com uma função de verossimilhança completamente especificada e, na ausência de um critério mais objetivo, a seleção do modelo deve levar isto em conta. Assim, para os dados do experimento com a linhagem TA98 expostas ao inseticida Fipronil, na ausência da fração microsomal S9, o modelo tradicional mais bem ajustado e com melhor propriedades estatísticas é aquele com preditor de Myers e componente aleatório BN-2.

Para os dados do experimento com a linhagem TA98, exposta ao inseticida Fipronil e na presença de S9, os modelos pré-selecionados levando em consideração os valores de $\Delta^{(AIC)}$ e de bondade de ajuste (Tabela 14) foram aqueles com preditores de Krewski, Margolin e Myers, todos com componente aleatório Poisson, que está de acordo com os resultados obtidos pelo teste *score* de Dean para superdispersão (Tabela 9). Uma vez que a análise de resíduos e diagnósticos desses modelos não revelaram nenhuma informação relevante que favorecesse um ou outro (ver Figuras 28 a 30), pode-se utilizar os mesmos argumentos de conveniência estatística, isto é, normalidade das distribuições amostrais

dos estimadores dos parâmetros e verossimilhança completamente especificada. Assim, deve-se selecionar o modelo com preditor de Myers e componente aleatório Poisson.

Tratando-se dos modelos ajustados aos dados do experimento com a linhagem TA98, exposta ao inseticida Tiametoxam na ausência de S9, os modelos pré-selecionados por $\Delta^{(AIC)}$ (Tabela 15) não estreitam a busca pelo modelo mais adequado, visto que esses critérios apenas excluíram os modelos com componente aleatório quase-potência e que, como visto, o teste escore de Dean não rejeitou a hipótese de equidispersão.

Assim, deve-se proceder com a análise de resíduos e diagnósticos dos modelos com preditor de Bernstein e componente aleatório Poisson (Figura 31) e com preditor de Stead, também com componente aleatório Poisson (Figura 32). O comportamento dos componentes de *deviance versus* doses apresentou o padrão nulo em ambos os modelos. Este padrão foi também observado no gráfico de r_i^{d*} *versus* $\hat{\eta}_i$ apenas no modelo com preditor de Stead, pois há ligeira mudança de amplitude neste gráfico para o modelo com preditor de Bernstein. É necessário, no entanto, considerar que as medidas de influência e curvatura locais excederam os valores nominais em um maior número de observações no modelo com preditor de Stead, que, além disso, apresentou os piores desempenhos na normalidade da distribuições amostrais dos estimadores. Por isto, e levando em consideração as boas e amplamente conhecidas propriedades dos MLGs, o fato de que o teste para bondade de ajuste rejeitou a hipótese de falta de ajuste para o modelo com preditor de Bernstein, e pelo princípio da parcimônia, o modelo selecionado para os dados do experimento com a linhagem TA98, exposta ao inseticida Tiametoxam na ausência de S9 deve ser aquele com preditor de Bernstein e componente aleatório Poisson.

Opositivamente, ao se considerar os dados provenientes dessa linhagem, exposta ao mesmo inseticida, mas na presença de S9, o critério $\Delta^{(AIC)}$ (Tabela 16) indica que todos os modelos com o preditor de Bernstein devem ser descartados, além de apresentar falta de ajuste e superdispersão. Assim, os modelos pré-selecionados, para os quais devem ser obtidos os gráficos de resíduos e de medidas de diagnósticos são aqueles com preditor de Stead (Figura 33), tanto com componente aleatório BN-2 quanto com quase-Poisson (Figura 34).

Uma vez que ambos os modelos possuem a mesma função para descrever a média do número de revertentes, o preditor de Stead, os gráficos de componentes de *deviance versus* doses e *versus* $\hat{\eta}_i$ apresentam o mesmo padrão – neste caso, o padrão nulo. O gráfico de resíduos ordenados com envelope simulado *versus* quantis da distribuição normal padrão revelam, em ambos os modelos, distribuições mais assimétricas do que aquelas atribuídas aos componentes aleatórios e observações fora dos envelopes.

É claro que, no caso do modelo com componente aleatório quase-Poisson, os envelopes foram gerados a partir da PG, que pode não retratar de fato o processo de geração dos dados, pois nenhuma distribuição completa foi especificada. Nesta linha de pensamento, outras distribuições poderiam ter sido usadas para gerar os envelopes, como a mistura

Poisson-Gaussiana Inversa (DEAN; LAWLESS; WILLMOT, 1989), a BN-1 (GREENE, 2008) ou a Poisson-Tweedie (JØRGENSEN, 1997, p. 165), respeitando-se a forma da variância escolhida, $\phi\mu_i$.

Há, sob essa perspectiva, evidência contra o componente aleatório BN-2, mas isto não constitui evidência a favor do componente aleatório quase-Poisson. De fato, mesmo as medidas de influência e curvatura locais são obtidas apenas de forma aproximada para funções de quase-verossimilhança e, nestes quesitos, os desempenhos dos dois modelos foram igualmente bons. Assim, fazendo uso do último critério para a seleção, que é o de verossimilhança completamente especificada conforme estabelecido na seção 4.4, o modelo escolhido, dentre os tradicionais, para os dados da linhagem TA98 exposta ao inseticida Tiametoxam na presença de S9, foi aquele com preditor de Stead e componente aleatório BN-2.

Já para os experimentos com a linhagem TA100 de *S. typhimurium*, os dois modelos pré-selecionados ajustados aos dados provenientes da exposição ao Fipronil na ausência de S9, segundo os critérios de $\Delta^{(AIC)}$ e de bondade de ajuste (Tabela 17), foram aqueles com preditor de Margolin e com preditor de Myers, ambos com componente aleatório BN-2.

Os gráficos de resíduos e de medidas de diagnóstico apresentados respectivamente nas Figuras 35 e 36 para os modelos com preditor de Margolin e de Myers são muito próximos, e, a partir deles, é impossível decidir sobre o melhor ajuste. Assim, devido as melhores propriedades das distribuições amostrais dos estimadores dos parâmetros de modelos utilizando o preditor de Myers, este deve ser selecionado. É necessário notar, entretanto, que trata-se de uma escolha difícil, entre dois ruins, visto que os gráficos de r_i^{d*} versus doses apresentaram clara tendência sistemática, possivelmente devido à ausência de um termo polinomial na dose na função que descreve a ação mutagênica da substância. Myers et al. (1981) já haviam alertado sobre esses possíveis comportamentos complexos e também consideraram a inclusão de potências nas doses em seus preditores não-lineares, mas foram, de forma geral, ignorados na literatura da área. Além disso, notaram-se observações fora do envelope simulado e localmente influentes fora do controle negativo.

Os problemas supracitados foram também encontrados nos modelos pré-selecionados para os dados da exposição da mesma linhagem a esse inseticida na presença de S9 (Tabela 18), cujos gráficos de resíduos e diagnósticos estão apresentados nas Figuras 37 a 40, que se referem, respectivamente, aos modelos com preditor de Margolin e componente aleatório BN-2, preditor de Margolin e componente aleatório quase-potência, preditor de Myers e componente aleatório BN-2, e preditor de Myers e componente aleatório quase-potência.

Para os modelos pré-selecionados com componente aleatório quase-potência, notam-se que os valores absolutos dos resíduos são muito baixos e isto se deveu ao uso de componentes de *deviance* não-escalonada. Vista a notável amplitude dos envelopes nos gráficos de resíduos ordenados para os modelos com componente aleatório quase-potência, pode-se entender que o uso da distribuição Tweedie discretizada para gerar essas estatísticas é

insatisfatório, trazendo pouca ou nenhuma informação sobre a distribuição que gerou os dados, sobretudo nas caudas. Também para a família quase-potência, assim como para a quase-Poisson, as medidas de diagnóstico, influência e curvatura locais, são apenas aproximadas e podem estar comprometidas, da mesma forma que os envelopes simulados.

Assim, desse conjunto de modelos pré-selecionados, utilizando os critérios de melhor condições na normalidade assintótica e de verossimilhança completamente especificada, deve-se selecionar o modelo com preditor de Myers e componente aleatório BN-2. Outro argumento para a escolha desse modelo está baseado na observação do gráfico de dispersão na Figura 22 e no modelo já selecionado para os dados da linhagem TA100 exposta ao Fipronil na ausência de S9: não há razão para supor que nem o processo de geração dos dados e tampouco a forma sistemática do modelo sejam diferentes na presença e na ausência de S9. Assim, como o modelo selecionado para a ausência de S9 é composto por preditor de Myers e componente aleatório BN-2, também deve ser selecionado um modelo com essas características para os dados obtidos na presença do SAM.

Os modelos tradicionais pré-selecionados ajustados aos dados de exposição da linhagem TA100 ao inseticida Tiametoxam na ausência de S9 foram, de acordo com os critérios de $\Delta^{(AIC)}$ e de bondade de ajuste (Tabela 19), aqueles compostos por preditor de Breslow e componente aleatório BN-2, por preditor de Stead e componente aleatório BN-2 e por preditor de Stead e componente aleatório quase-Poisson. Os gráficos de resíduos e diagnósticos apresentados, respectivamente, nas Figuras 41 a 43 revelaram padrão nulo e poucas observações com resíduos ordenados fora dos envelopes simulados, além de que os maiores valores das medidas de influência e curvatura local se deram em observações no controle negativo. Desta forma, nenhum dos três modelos pode ser imediatamente descartado e as propriedades das distribuições amostrais dos estimadores dos parâmetros podem ser levadas em consideração, apontando para o modelo com preditor de Breslow e componente aleatório BN-2.

Finalmente, os modelos pré-selecionados, ajustados aos dados dos experimentos com a linhagem TA100 exposta ao inseticida Tiametoxam na presença de S9, foram, segundo os critérios de $\Delta^{(AIC)}$ e de bondade de ajuste disponíveis na Tabela 20), aqueles com preditor de Breslow e componente aleatório BN-2, com preditor de Breslow e componente aleatório quase-potência, e com preditor de Margolin com componente aleatório quase-potência.

Os gráficos de resíduos e diagnósticos desses modelos estão apresentados respectivamente nas Figuras 44 a 46. Uma observação, localmente influente e de curvatura grande fora do controle negativo, apresentou resíduo muito maior que as demais em todos os três modelos e também fora dos envelopes simulados. Esta observação em particular é *outlier*, uma vez que, além de ponto inconsistente, apresenta *leverage* pequeno (ver, por exemplo, McCullagh e Nelder (1989, p. 406)).

Devido aos problemas nos envelopes simulados obtidos pela distribuição Tweedie discretizada, aos critérios previamente estabelecidos para seleção de modelos e pelo prin-

cípio da parcimônia, entende-se que o modelo que deve ser selecionado para esses dados é aquele composto por preditor de Breslow e componente aleatório BN-2.

Neste momento, é importante observar que, de acordo com a formulação tradicional, os modelos devem ser ajustados separadamente aos dados de cada linhagem, substância sob investigação e sistema de ativação metabólica, que, em alguns casos, implica em modelos com interações implícitas entre substância e sistema de ativação metabólica. Este tipo de formulação, além de pouco elegante, deixa evidente um problema grave, caracterizado pela impossibilidade de considerar modelos com intercepto comum. Independente do nível de S9, o número de revertentes espontâneos é o mesmo para as linhagens TA98 e TA100, de forma que o intercepto comum, ao menos com relação ao sistema de ativação metabólica, não é apenas desejável a fim da diminuição no número de parâmetros, mas faz parte da validação do experimento em si.

Considerando os conjuntos de dados obtidos no presente trabalho, a formulação tradicional implica em oito modelos finais. Para os dados dos experimentos com a linhagem TA98, os modelos tradicionais selecionados para o Fipronil, respectivamente para S9(−) e S9(+), são compostos por preditores de Myers e componentes aleatórios BN-2 e Poisson. Há pouca razão para supor que, se a forma sistemática é a mesma, os dados provenientes da exposição de uma mesma linhagem a um mesmo químico tenham sido gerados por processos aleatórios distintos. Além disso, as estimativas dos interceptos de cada um dos dois modelos é realizada de forma independente, apresentando a supracitada redundância, dos pontos de vista teórico e prático. Assim, é mais razoável supor que houve falha em detectar superdispersão nos dados obtidos na presença de S9, e, assim, elaborar um modelo de intercepto comum para S9(−) e S9(+). Isto pode ser feito no contexto dos MNLGVs e os resultados dessa abordagem encontram-se na seção 5.2.

Os modelos tradicionais selecionados para os dados dos experimentos com a linhagem TA98 exposta ao inseticida Tiametoxam indicam que há diferença nos mecanismos de ação tóxica e mutagênica, se na ausência (modelo com preditor de Bernstein e componente aleatório Poisson) ou se na presença de S9 (modelo com preditor de Stead e componente aleatório BN-2), de forma que faz pouco sentido supor que um mesmo processo aleatório tenha gerado os dados. Assim, nos casos como este, em que há mudança nos mecanismos de ação, a abordagem tradicional é suficientemente boa e possui como alternativa proposta neste estudo modelos de regressão utilizando a distribuição de Skellam, cujos resultados estão apresentados também na seção 5.2.

Já os modelos selecionados para os dados dos experimentos com a linhagem TA100 permitem visualizar mais claramente as possíveis falhas nessas interações implícitas: para o Fipronil, os modelos tradicionais selecionados são compostos por preditor de Myers e componente aleatório BN-2, tanto para S9(−) quanto para S9(+); e, similarmente para o Tiametoxam, os modelos tradicionais selecionados são compostos por preditor de Breslow e componente aleatório BN-2. Como no caso dos dados da linhagem TA98 exposta ao

inseticida Fipronil, faz sentido elaborar modelos com interações explícitas, buscando prioritariamente um modelo com intercepto comum, e considerar possíveis dependências das covariáveis, doses e fator S9, no parâmetro k da BN-2. Além disso, modelos de regressão Skellam também são outras alternativas para ajuste aos dados da linhagem TA100.

Figura 21 - Gráficos de dispersão dos dados observados nos experimentos com a linhagem TA98 no ensaio de *Salmonella*/microsoma. Os painéis superiores (A e B) se referem aos dados da exposição ao inseticida Tiametoxam e os painéis inferiores (C e D) aos dados da exposição ao Fipronil. À esquerda (A e C), os dados na presença da fração microsomal S9 e à direita (B e D), estão os dados na ausência de S9. Todas as observações da Tabela 5, em escala logarítmica, estão denotadas pelo símbolo o, as linhas sólidas representam a suavização dos dados e as linhas tracejadas as respectivas bandas de confiança a 95%

Figura 22 - Gráficos de dispersão dos dados observados nos experimentos com a linhagem TA100 no ensaio de *Salmonella*/microsoma. Os painéis superiores (A e B) se referem aos dados da exposição ao inseticida Tiametoxam e os painéis inferiores (C e D) aos dados da exposição ao Fipronil. À esquerda (A e C), os dados na ausência da fração microsomal S9 e à direita (B e D), estão os dados na presença de S9. Todas as observações da Tabela 6, em escala logarítmica, estão denotadas pelo símbolo ○, as linhas sólidas representam a suavização dos dados e as linhas tracejadas as respectivas bandas de confiança a 95%

Tabela 9 - Testes escore de Dean para superdispersão, de acordo com cada preditor tradicional, para os dados do experimento de *Salmonella*/microsossoma com a linhagem TA98, utilizando o inseticida Fipronil, na presença e na ausência da fração microsossomal S9 (sistema de ativação metabólica)

SAM	Preditor	V	S	p
S9(-)	Bernstein	$\phi\mu$	15,351	< 0,00001
S9(-)	Bernstein	$\mu(1 + k^{-1}\mu)$	5,902	< 0,00001
S9(-)	Krewski	$\phi\mu$	2,65	0,004 02
S9(-)	Krewski	$\mu(1 + k^{-1}\mu)$	2,926	0,001 72
S9(-)	Margolin	$\phi\mu$	2,64	0,004 14
S9(-)	Margolin	$\mu(1 + k^{-1}\mu)$	2,924	0,001 73
S9(-)	Myers	$\phi\mu$	2,641	0,004 14
S9(-)	Myers	$\mu(1 + k^{-1}\mu)$	2,924	0,001 73
S9(-)	Stead	$\phi\mu$	3,104	0,000 95
S9(-)	Stead	$\mu(1 + k^{-1}\mu)$	4,275	0,000 01
S9(+)	Bernstein	$\phi\mu$	1,568	0,058 44
S9(+)	Bernstein	$\mu(1 + k^{-1}\mu)$	1,658	0,048 68
S9(+)	Krewski	$\phi\mu$	-1,007	0,843 07
S9(+)	Krewski	$\mu(1 + k^{-1}\mu)$	1,109	0,1338
S9(+)	Margolin	$\phi\mu$	-1,012	0,844 24
S9(+)	Margolin	$\mu(1 + k^{-1}\mu)$	1,108	0,133 98
S9(+)	Myers	$\phi\mu$	-1,012	0,8442
S9(+)	Myers	$\mu(1 + k^{-1}\mu)$	1,108	0,133 97

Tabela 10 - Testes escore de Dean para superdispersão, de acordo com cada preditor tradicional, para os dados do experimento de *Salmonella*/microsossoma com a linhagem TA98, utilizando o inseticida Tiametoxam, na presença e na ausência da fração microsossomal S9 (sistema de ativação metabólica)

SAM	Preditor	V	S	p
S9(-)	Bernstein	$\phi\mu$	1,605	0,054 23
S9(-)	Bernstein	$\mu(1 + k^{-1}\mu)$	1,544	0,061 31
S9(-)	Stead	$\phi\mu$	0,994	0,1602
S9(-)	Stead	$\mu(1 + k^{-1}\mu)$	0,973	0,165 21
S9(+)	Bernstein	$\phi\mu$	5,115	< 0,00001
S9(+)	Bernstein	$\mu(1 + k^{-1}\mu)$	4,925	< 0,00001
S9(+)	Stead	$\phi\mu$	1,907	0,028 24
S9(+)	Stead	$\mu(1 + k^{-1}\mu)$	1,848	0,032 28

Tabela 11 - Testes escore de Dean para superdispersão, de acordo com cada preditor tradicional, para os dados do experimento de *Salmonella*/microsoma com a linhagem TA100, utilizando o inseticida Fipronil, na presença e na ausência da fração microsomal S9 (sistema de ativação metabólica)

SAM	Preditor	V	S	p
S9(-)	Bernstein	$\phi\mu$	58,687	$< 0,00001$
S9(-)	Bernstein	$\mu(1 + k^{-1}\mu)$	74,816	$< 0,00001$
S9(-)	Margolin	$\phi\mu$	36,761	$< 0,00001$
S9(-)	Margolin	$\mu(1 + k^{-1}\mu)$	48,058	$< 0,00001$
S9(-)	Myers	$\phi\mu$	36,724	$< 0,00001$
S9(-)	Myers	$\mu(1 + k^{-1}\mu)$	48,029	$< 0,00001$
S9(+)	Bernstein	$\phi\mu$	207,953	$< 0,00001$
S9(+)	Bernstein	$\mu(1 + k^{-1}\mu)$	285,764	$< 0,00001$
S9(+)	Margolin	$\phi\mu$	50,367	$< 0,00001$
S9(+)	Margolin	$\mu(1 + k^{-1}\mu)$	82,442	$< 0,00001$
S9(+)	Myers	$\phi\mu$	50,008	$< 0,00001$
S9(+)	Myers	$\mu(1 + k^{-1}\mu)$	81,913	$< 0,00001$

Tabela 12 - Testes escore de Dean para superdispersão, de acordo com cada preditor tradicional, para os dados do experimento de *Salmonella*/microsoma com a linhagem TA100, utilizando o inseticida Tiametoxam, na presença e na ausência da fração microsomal S9 (sistema de ativação metabólica)

SAM	Preditor	V	S	p
S9(-)	Breslow	$\phi\mu$	59,633	$< 0,00001$
S9(-)	Breslow	$\mu(1 + k^{-1}\mu)$	59,677	$< 0,00001$
S9(-)	Krewski	$\phi\mu$	65,6	$< 0,00001$
S9(-)	Krewski	$\mu(1 + k^{-1}\mu)$	68,253	$< 0,00001$
S9(-)	Margolin	$\phi\mu$	59,757	$< 0,00001$
S9(-)	Margolin	$\mu(1 + k^{-1}\mu)$	57,537	$< 0,00001$
S9(-)	Myers	$\phi\mu$	59,943	$< 0,00001$
S9(-)	Myers	$\mu(1 + k^{-1}\mu)$	57,903	$< 0,00001$
S9(-)	Stead	$\phi\mu$	53,909	$< 0,00001$
S9(-)	Stead	$\mu(1 + k^{-1}\mu)$	53,074	$< 0,00001$
S9(+)	Breslow	$\phi\mu$	32,097	$< 0,00001$
S9(+)	Breslow	$\mu(1 + k^{-1}\mu)$	33,922	$< 0,00001$
S9(+)	Krewski	$\phi\mu$	35,412	$< 0,00001$
S9(+)	Krewski	$\mu(1 + k^{-1}\mu)$	38,047	$< 0,00001$
S9(+)	Margolin	$\phi\mu$	32,839	$< 0,00001$
S9(+)	Margolin	$\mu(1 + k^{-1}\mu)$	34,854	$< 0,00001$
S9(+)	Myers	$\phi\mu$	32,912	$< 0,00001$
S9(+)	Myers	$\mu(1 + k^{-1}\mu)$	34,953	$< 0,00001$
S9(+)	Stead	$\phi\mu$	31,085	$< 0,00001$
S9(+)	Stead	$\mu(1 + k^{-1}\mu)$	32,764	$< 0,00001$

Tabela 13 - Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de *Salmonella*/microsossoma utilizando o inseticida Fipronil na ausência da fração microsossomal S9(-), com a linhagem TA98

CA	Preditor	AIC _c	$\Delta^{(AIC)}$	X^2	$\tilde{D}$	GL
Poisson	Bernstein	255,51	73,01	126,76	114,837	21
Poisson	Krewski	184,67	2,18	40,969	41,340	20
Poisson	Margolin	184,58	2,08	40,902	41,244	20
Poisson	Myers	184,58	2,09	40,905	41,248	20
Poisson	Stead	190,08	7,59	44,188	43,788	19
BN-2	Bernstein	204,56	22,06	27,231	23,013	20
BN-2	Krewski	182,57	0,07	25,313	26,621	19
BN-2	Margolin	187,54	5,04	40,901	41,244	19
BN-2	Myers	182,49	0	25,319	25,601	19
BN-2	Stead	186,63	4,14	24,058	23,827	18
Quase-Poisson	Bernstein	205,09	22,59	—	19,025	20
Quase-Poisson	Krewski	183,37	0,87	—	20,181	19
Quase-Poisson	Margolin	183,31	0,82	—	20,167	19
Quase-Poisson	Myers	183,31	0,82	—	20,168	19
Quase-Poisson	Stead	188,09	5,60	—	18,828	18
Quase-potência	Bernstein	207,92	25,43	—	18,715	19
Quase-potência	Krewski	185,83	3,33	—	20,280	18
Quase-potência	Margolin	185,76	3,27	—	20,261	18
Quase-potência	Myers	185,77	3,27	—	20,262	18
Quase-potência	Stead	189,32	6,82	—	19,047	17

Figura 23 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Krewski e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Fipronil, na ausência da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 24 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Krewski e componente aleatório quase-Poisson, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Fipronil, na ausência da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 25 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Margolin e componente aleatório quase-Poisson, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Fipronil, na ausência da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudantizados; $r_{(i)}^{d*}$: componentes de *deviance* estudantizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 26 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Myers e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Fipronil, na ausência da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 27 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Myers e componente aleatório quase-Poisson, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Fipronil, na ausência da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudantizados; $r_{(i)}^{d*}$: componentes de *deviance* estudantizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Tabela 14 - Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de *Salmonella*/microsossoma utilizando o inseticida Fipronil na presença da fração microsossomal S9(+), com a linhagem TA98

CA	Preditor	AIC _c	$\Delta^{(AIC)}$	X^2	$\tilde{D}$	GL
Poisson	Bernstein	168,36	13,66	33,437	32,340	21
Poisson	Krewski	154,74	0,04	16,167	16,053	20
Poisson	Margolin	154,70	0	16,134	16,017	20
Poisson	Myers	154,70	0	16,135	16,018	20
BN-2	Bernstein	169,05	14,35	25,359	24,285	20
BN-2	Krewski	157,11	2,41	13,485	13,379	19
BN-2	Margolin	157,66	2,96	16,134	16,017	19
BN-2	Myers	157,07	2,37	13,457	13,349	19
Quase-Poisson	Bernstein	170,23	15,53	–	20,311	20
Quase-Poisson	Krewski	156,77	2,06	–	19,860	19
Quase-Poisson	Margolin	156,71	2,01	–	19,855	19
Quase-Poisson	Myers	156,72	2,02	–	19,855	19
Quase-potência	Bernstein	172,0	17,30	–	19,994	19
Quase-potência	Krewski	156,43	1,72	–	19,844	18
Quase-potência	Margolin	156,35	1,65	–	19,831	18
Quase-potência	Myers	156,35	1,65	–	19,831	18

Figura 28 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Krewski e componente aleatório Poisson, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Fipronil, na presença da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudantizados; $r_{(i)}^{d*}$: componentes de *deviance* estudantizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 29 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Margolin e componente aleatório Poisson, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Fipronil, na presença da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 30 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Myers e componente aleatório Poisson, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Fipronil, na presença da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudantizados; $r_{(i)}^{d*}$: componentes de *deviance* estudantizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Tabela 15 - Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de *Salmonella*/microsoma utilizando o inseticida Tiametoxam na ausência da fração microsomal S9(–), com a linhagem TA98

CA	Preditor	AIC _c	$\Delta^{(AIC)}$	X^2	$\tilde{D}$	GL
Poisson	Bernstein	245,61	0,13	50,811	49,667	35
Poisson	Stead	245,48	0	45,552	44,638	33
BN-2	Bernstein	246,08	0,6	37,751	36,738	34
BN-2	Stead	247,36	1,88	37,415	36,556	32
Quase-Poisson	Bernstein	247,18	1,71	–	34,213	34
Quase-Poisson	Stead	248,56	3,08	–	32,338	32
Quase-potência	Bernstein	249,70	4,22	–	34,169	33
Quase-potência	Stead	251,37	5,90	–	32,253	31

Figura 31 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Bernstein e componente aleatório Poisson, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Tiametoxam, na ausência da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 32 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Stead e componente aleatório Poisson, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Tiametoxam, na ausência da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Tabela 16 - Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de *Salmonella*/microsoma utilizando o inseticida Tiametoxam na presença da fração microsomal S9(+), com a linhagem TA98

CA	Preditor	AIC _c	$\Delta^{(AIC)}$	X^2	D	GL
Poisson	Bernstein	297,27	23,55	87,324	88,493	39
Poisson	Stead	274,12	0,4	58,197	60,549	37
BN-2	Bernstein	284,23	10,51	40,630	41,876	38
BN-2	Stead	273,72	0	40,274	42,469	36
Quase-Poisson	Bernstein	285,62	11,9	—	39,552	38
Quase-Poisson	Stead	274,67	0,95	—	38,495	36
Quase-potência	Bernstein	286,93	13,21	—	39,000	37
Quase-potência	Stead	276,37	2,65	—	37,000	35

Figura 33 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Stead e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de *Salmonella*/microsossoma (TA98) realizado com o inseticida Tiametoxam, na presença da fração microsossoma S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 34 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Stead e componente aleatório quase-Poisson, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Tiametoxam, na presença da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Tabela 17 - Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de *Salmonella*/microsossoma utilizando o inseticida Fipronil na ausência da fração microsossomal S9(–), com a linhagem TA100

CA	Preditor	AIC _c	$\Delta^{(AIC)}$	X^2	D	GL
Poisson	Bernstein	773,06	399,43	525,468	515,950	33
Poisson	Margolin	615,65	242,03	342,636	356,144	32
Poisson	Myers	615,33	241,71	342,326	355,825	32
BN-2	Bernstein	777,01	403,38	525,468	515,915	32
BN-2	Margolin	373,65	0,02	31,796	35,592	31
BN-2	Myers	373,63	0	31,795	35,594	31
Quase-Poisson	Bernstein	390,18	16,56	–	32,402	32
Quase-Poisson	Margolin	378,98	5,35	–	33,262	31
Quase-Poisson	Myers	378,94	5,32	–	33,262	31
Quase-potência	Bernstein	380,22	6,59	–	36,225	31
Quase-potência	Margolin	377,23	3,61	–	35,065	30
Quase-potência	Myers	377,20	3,58	–	35,067	30

Figura 35 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Margolin e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Fipronil, na ausência da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 36 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Myers e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Fipronil, na ausência da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Tabela 18 - Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de *Salmonella*/microsossoma utilizando o inseticida Fipronil na presença da fração microssomal S9(+), com a linhagem TA100

CA	Preditor	AIC _c	$\Delta^{(AIC)}$	X^2	D	GL
Poisson	Bernstein	1840,37	1426,95	1898,538	1555,326	38
Poisson	Margolin	796,41	382,99	490,868	509,026	37
Poisson	Myers	792,49	379,07	487,655	505,109	37
BN-2	Bernstein	433,91	20,49	47,823	38,916	37
BN-2	Margolin	413,63	0,21	36,843	39,250	36
BN-2	Myers	413,42	0	36,854	39,258	36
Quase-Poisson	Bernstein	478,15	64,73	—	31,130	37
Quase-Poisson	Margolin	432,21	18,79	—	38,369	36
Quase-Poisson	Myers	431,89	18,47	—	38,324	36
Quase-potência	Bernstein	418,67	5,25	—	39,937	36
Quase-potência	Margolin	414,06	0,64	—	39,324	35
Quase-potência	Myers	414,46	1,04	—	34,706	35

Figura 37 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Margolin e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Fipronil, na presença da fração microssomal S9. r_i^{d*} : componentes de *deviance* estudantizados; $r_{(i)}^{d*}$: componentes de *deviance* estudantizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 38 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Margolin e componente aleatório quase-potência, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Fipronil, na presença da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudantizados; $r_{(i)}^{d*}$: componentes de *deviance* estudantizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 39 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Myers e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Fipronil, na presença da fração microssomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 40 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Myers e componente aleatório quase-potência, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Fipronil, na presença da fração microssomal S9. r_i^{d*} : componentes de *deviance* estudantizados; $r_{(i)}^{d*}$: componentes de *deviance* estudantizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Tabela 19 - Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de *Salmonella*/microsossoma utilizando o inseticida Tiametoxam na ausência da fração microssomal S9(-), com a linhagem TA100

CA	Preditor	AIC _c	$\Delta^{(AIC)}$	X^2	D	GL
Poisson	Breslow	804,45	432,46	534,103	566,771	32
Poisson	Krewski	863,35	491,35	584,090	625,670	32
Poisson	Margolin	801,12	429,13	535,479	563,443	32
Poisson	Myers	803,28	431,29	537,025	565,606	32
Poisson	Stead	737,55	365,56	486,119	497,314	31
BN-2	Breslow	373,9	1,9	30,137	35,880	31
BN-2	Krewski	376,54	4,55	29,369	35,885	31
BN-2	Margolin	374,13	2,13	30,067	35,895	31
BN-2	Myers	374,21	2,22	30,033	35,894	31
BN-2	Stead	371,99	0	32,475	35,846	30
Quase-Poisson	Breslow	375,09	3,1	—	33,957	31
Quase-Poisson	Krewski	378,69	6,69	—	34,278	31
Quase-Poisson	Margolin	374,63	2,64	—	33,671	31
Quase-Poisson	Myers	374,78	2,79	—	33,703	31
Quase-Poisson	Stead	373,15	1,16	—	41,714	30
Quase-potência	Breslow	375,58	3,59	—	32,000	30
Quase-potência	Krewski	374,0	2,01	—	32,000	30
Quase-potência	Margolin	374,44	2,45	—	31,995	30
Quase-potência	Myers	374,47	2,47	—	32,000	30
Quase-potência	Stead	375,94	3,95	—	31,364	29

Figura 41 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Breslow e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 42 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Stead e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 43 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Stead e componente aleatório quase-Poisson, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Tabela 20 - Critérios de informação e de bondade de ajuste dos modelos com diferentes componentes aleatórios e preditores tradicionais ajustados aos dados oriundos do ensaio de *Salmonella*/microsossoma utilizando o inseticida Tiametoxam na presença da fração microsossomal S9(+), com a linhagem TA100

CA	Preditor	AIC _c	$\Delta^{(AIC)}$	X^2	D	GL
Poisson	Breslow	578,11	178,66	331,689	295,512	38
Poisson	Krewski	607,75	208,3	361,702	325,15	38
Poisson	Margolin	586,17	186,72	338,477	303,573	38
Poisson	Myers	586,8	187,35	339,128	304,201	38
Poisson	Stead	569,98	170,53	322,533	284,913	37
BN-2	Breslow	401,14	1,69	49,371	40,333	37
BN-2	Krewski	404,75	5,3	48,502	40,402	37
BN-2	Margolin	402,27	2,82	48,771	40,367	37
BN-2	Myers	402,34	2,89	48,76	40,368	37
BN-2	Stead	402,04	2,59	49,922	40,24	36
Quase-Poisson	Breslow	408,65	9,2	—	33,855	37
Quase-Poisson	Krewski	412,6	13,15	—	34,16	37
Quase-Poisson	Margolin	409,67	10,22	—	34,081	37
Quase-Poisson	Myers	409,76	10,31	—	34,086	37
Quase-Poisson	Stead	409,98	10,53	—	42,584	36
Quase-potência	Breslow	399,45	0	—	36,531	36
Quase-potência	Krewski	402,8	3,35	—	38,717	36
Quase-potência	Margolin	400,61	1,16	—	37,495	36
Quase-potência	Myers	409,05	9,6	—	19,937	36
Quase-potência	Stead	402,41	2,96	—	29,601	35

Figura 44 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Breslow e componente aleatório binomial negativo tipo-2, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Tiametoxam, na presença da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 45 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Breslow e componente aleatório quase-potência, pré-selecionado para os dados do ensaio de *Salmonella*/microsossoma (TA100) realizado com o inseticida Tiametoxam, na presença da fração microssomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 46 - Resíduos e diagnósticos para o MNLG tradicional com preditor de Margolin e componente aleatório quase-potência, pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Tiametoxam, na presença da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

5.2 Modelos propostos

Nesta seção são apresentados os resultados e discussões sobre os MNLGVs da família BN-2 e os MLGVs com componente aleatório Skellam, nesta ordem. Assim como na seção 5.1.2, primeiro os resultados para a linhagem TA98 e depois para a linhagem TA100 e, em cada linhagem, primeiro para o Fipronil e depois para o Tiametoxam. Em todos os casos, a ordem de apresentação respeita os passos para seleção de modelos estabelecidos na seção 4.4. Não há ordem na apresentação, contudo, para os dados na ausência e na presença de S9, uma vez que na presente seção, trata-se da integração desses dados, em modelos com interação explícita. Note que esta é uma seção de seleção dos modelos finais para cada conjunto de dados, de forma que os resultados de inferência e medidas de interesse estão apresentados na seção 5.3.

Para os dados do experimento realizado com a linhagem TA98 exposta a concentrações do inseticida Fipronil, o modelo final selecionado foi o Modelo III, cujos preditor e componente aleatório constam na Tabela 21. A seleção se deu por testes de razão de verossimilhanças, partindo do Modelo I (Tabela 21), que, não fosse pelo componente aleatório Poisson do MNLG tradicional selecionado para os dados na presença de S9, seria idêntico a um MNLG incluídas todas interações explícitas do fator S9 com todos os parâmetros. Os gráficos de resíduos e diagnósticos do modelo selecionado estão disponíveis na Figura 47, apresentando todas as características desejáveis.

O modelo selecionado é um MNLG com interação do SAM S9 nos parâmetros de mutagênese e de toxicidade. Os demais modelos intermediários, que foram descartados também por testes de razão de verossimilhanças, incluindo aqueles com termos polinomiais no preditor do parâmetro k da BN-2, são MNLGVs estritos. Assim, a classe de MNLGVs se fez útil, fornecendo um contexto dentro do qual o processo de seleção pôde fazer uso de critérios objetivos e fortes, como os testes de razão de verossimilhanças para modelos aninhados.

Ainda sobre a linhagem TA98 de *S. typhimurium*, o experimento realizado com o inseticida Tiametoxam revelou que os mecanismos de ação, na presença e na ausência do SAM S9, da mutagenicidade e da toxicidade são distintos tanto com relação à presença de superdispersão quanto com à forma do componente sistemático dos modelos (ver os modelos pré-selecionados para esses dados na seção 5.1.2). Assim especificar um MNLGV com tantas restrições, ainda que possível, é redundante do ponto de vista estatístico e pouco esclarecedor do ponto de vista biológico e, por isso, nenhum desenvolvimento utilizando modelos não-lineares vetoriais da família binomial negativa tipo-2 foi feito para esses experimentos.

Já para os experimentos com a linhagem TA100, o MNLGV, com componente aleatório BN-2, para os dados do experimento com exposição ao Fipronil selecionado foi o Modelo III da Tabela 22 e, para os dados com exposição ao inseticida Tiametoxam, foi

selecionado o Modelo II, na Tabela 23.

Do modelo selecionado para o Fipronil (Modelo III da Tabela 22) é interessante notar que não há qualquer efeito do fator S9, nem sobre os parâmetros do preditor da média BN-2 e nem sobre o preditor de apenas intercepto para o parâmetro k .

Esse modelo equivale a um MNLG tradicional, com preditor de Myers e componente aleatório BN-2, ajustado ao *pool* de dados, ignorando-se os níveis do SAM S9, que caracteriza um tipo de ajuste pouco comum na literatura da área (JEONG; FREEMAN; CLAXTON, 2010; SUEIRO; GARRIDO; ARAUJO, 2011; UMBUZEIRO et al., 2014; WARREN et al., 2015). Esta abordagem é geralmente evitada pela dificuldade de estabelecer modelos intermediários, algo solucionado no presente trabalho via MNLGVs; pela ausência de plataformas computacionais específicas para o ajuste de modelos aos dados do TA; e, principalmente, pela ingenuidade das análises estatísticas propostas nos protocolos para o desenvolvimento do ensaio de *Salmonella*/microsoma.

Os gráficos de resíduos e de medidas de diagnóstico para esse modelo estão apresentadas na Figura 48. Há uma tendência sistemática nos componentes de *deviance* estudada *versus* doses, que se manteve mesmo com inclusão de um termo quadrático na função do preditor que descreve a ação mutagênica, o que compromete a capacidade preditiva do modelo. Em contrapartida, o componente aleatório escolhido parece estar bem especificado, visto que a vasta maioria dos resíduos ordenados estão dentro do envelope simulado. Quanto às medidas de diagnósticos, apenas as observações no controle negativo apresentaram *leverage*, influência e curvatura locais acima dos valores nominais.

Sobre o modelo selecionado para os dados de exposição da linhagem TA100 ao Tiametoxam (Modelo II da Tabela 23), é interessante notar que não houve efeito do SAM S9 sobre a média do número de revertentes, de forma que os mecanismos de ação mutagênica e tóxica permanecem inalterados, mesmo após a metabolização da substância.

O efeito do fator S9 emerge, contudo, sobre o parâmetro de superdispersão, k , da BN-2, estabelecendo uma espécie de bloco, e cuja importância deve ser refletida sobre as distribuições preditivas. Outra hipótese, geralmente não avaliada, é que a metabolização da substância exerce efeito, não sobre a média, mas sobre a variabilidade da variável resposta – o número de revertentes por placa – e, se for o caso, medidas de segurança e políticas públicas precisam levar esse tipo de efeito em consideração. Note que há pouca ou nenhuma informação para defender essa afirmação, e novos experimentos controlados são necessários para verificar se existem evidências para corroborá-la.

Ainda sobre esse modelo selecionado, vale salientar que se trata de um MNLGV estrito, com um preditor não-linear para média e um preditor linear para o parâmetro de superdispersão. É certo que nos trabalhos de Cordeiro e Demétrio (2008), Cordeiro e Udo (2008) e Cordeiro, Previdelli e Samohyl (2008) foram considerados preditores para parâmetros de superdispersão, inculindo métodos de quase-verossimilhança para a estimação. Entretanto, a classe de modelos utilizada no presente trabalho não tem a restrição de que

o componente aleatório pertence necessariamente à família exponencial e, tratando-se dos dados do TA, a formulação dos MNLGVs é ampla e tem potencial para fornecer bons ajustes a dados de novos experimentos, unido as abordagens empírica e mecanicista, como a de Eckardt e Haynes (1977).

Podem-se observar, na Figura 49, os gráficos de resíduos e medidas de diagnóstico para o preditor da média do modelo selecionado. Nos painéis de resíduos de Pearson estudatizados, nota-se uma observação cujo valor de residuo é aproximadamente 8, muito maior que o das demais. Ao verificar o gráfico de resíduos ordenados também foi identificado que o resíduo dessa observação está fora do envelope simulado. Contudo, sua medida de *leverage* não ultrapassa o valor nominal, caracterizando um *outlier*. Com relação às medidas de influência e curvatura locais, apenas observações no controle negativo apresentaram valores altos, ao menos se tratando dos parâmetros associados à média da distribuição BN-2. Note que esses gráficos não foram obtidos para o parâmetro de superdispersão, uma vez que os padrões nulos para os resíduos, e tampouco os valores nominais para as medidas de diagnósticos, não estão bem estabelecidos para esse tipo de parâmetro, no contexto dos MNLGVs, que, certamente, merecem atenção em estudos teóricos.

Os MLGVs com componente aleatório Skellam não puderam ser ajustados aos dados de exposição ao Fipronil, tanto nos experimentos com a linhagem TA98 quanto com a linhagem TA100 de *S. typhimurium*, devido à falta de convergência, que, por sua vez, foi causada pela instabilidade numérica das funções modificadas de Bessel, presentes na distribuição, conforme a Equação (49).

Apesar dos esforços realizados neste trabalho para melhorar as funções disponíveis tanto na base do *software* R (R CORE TEAM, 2015) quanto aquelas presentes no pacote “Bessel” (MAECHLER, 2015), ainda não há precisão suficiente nas funções interdependentes modificadas, constantes no pacote “ames”, `dske1lam2()` e `Besse1Is2()`, para a análise de determinados conjuntos de dados, sobretudo nas situações em que os valores absolutos das observações são muito maiores que os preditores de μ_1 ou de μ_2 da distribuição de Skellam.

De fato, é notável que o efeito tóxico, e, por consequência, seu valor esperado (μ_2), é consideravelmente menor nos dados obtidos nos experimentos com o Fipronil. Alzaid e Omair (2010) propuseram uma outra formulação da distribuição de Skellam, que não depende da função modificada de Bessel de primeira espécie; ao invés, os autores utilizaram funções hipergeométricas regularizadas (ver Exemplos 13 e 17 de Abramowitz e Stegun (1972)). Esta abordagem pode ser eficaz para obter com sucesso estimativas dos parâmetros de um MLGV com componente aleatório Skellam, ainda que a implementação computacional seja inconveniente, principalmente em linguagem R: os pacotes “hypergeo” (HANKIN, 2015; HANKIN; LEE, 2006) e “fAsianOptions” (WUERTZ, 2013) trazem aproximações numéricas de funções das funções utilizadas por Alzaid e Omair (2010) e executar as transformações necessárias é computacionalmente intensivo. Estas extensões

poderão ser disponibilizadas em versões futuras do pacote “ames” e dados como os obtidos no presente trabalho para o inseticida Fipronil poderão ser analisados.

Como discutido na seção 5.1.2, há evidências para supor diferentes mecanismos de ação para a mutagenicidade e para a toxicidade quando na presença ou na ausência do SAM S9 para os dados obtidos nos experimentos com a linhagem TA98 exposta ao inseticida Tiametoxam. Assim, os modelos com componente aleatório Skellam foram ajustados separadamente para cada nível desse fator.

Para S9(−) não foi obtida nenhuma melhora no ajuste que favorecesse o MLGV no lugar do MLG previamente selecionado para esses dados na seção 5.1.2 (preditor de Bernstein e componente aleatório Poisson). Já para S9(+), o Modelo II da Tabela 24 apresentou melhor ajuste que o modelo tradicional com preditor de Stead e componente aleatório BN-2, visto a diferença de 12,58 no valor de AIC_c . Essencialmente, não há evidência empírica para o modelo tradicional e, além disso, o modelo proposto apresentou melhor condicionamento nas medidas de resíduos e diagnósticos (Figura 50).

O Modelo II da Tabela 25 foi selecionado, por testes de razão de verossimilhanças, dentre os MLGVs com componente aleatório Skellam para os dados da linhagem TA100 exposta ao inseticida Tiametoxam. Esse modelo apresenta interações significativas entre o SAM S9 e as doses. Na Figura 51 é possível observar que os resíduos padronizados, r_i^z , não apresentam o padrão nulo e que a maior parte dos resíduos ordenados não se encontra dentro do envelope simulado. Além disso, há diversas observações, fora do controle negativo, localmente influentes e com curvatura maior que o valor nominal.

Comparando-se via AIC_c os Modelos II da Tabela 23 com o Modelo II da Tabela 25, que são respectivamente iguais a 772.18 e a 889.68, pode-se entender que o melhor modelo para esse conjunto de dados é o MNLGV com preditor de Breslow para a média e preditor com efeito do fator S9 para o parâmetro de superdispersão da BN-2.

Neste momento, tendo em mãos os modelos finais selecionados para cada conjunto de dados, é interessante realizar as inferências e predições, e obter as medidas de significado biológico. Todas essas informações estão apresentadas e discutidas na seção 5.3.

Tabela 21 - Testes de razão de verossimilhanças para os MNGLVs encaixados da família binomial negativa tipo-2 (BN-2) com preditor de Myers ajustados aos dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Fipronil, na ausência e na presença da fração microsomal S9

Modelo	η_{1ij}	η_{2ij}	GL	$-L$	χ^2	p
I	$(\beta_{10j} + \beta_{11j}x_i) \exp(-\gamma_{1j}x_i)$	θ_{2j}	38	159,6604	—	—
II	$(\beta_{10} + \beta_{11j}x_i) \exp(-\gamma_{1j}x_i)$	θ_{2j}	39	163,395	3,735	0,053
III	$(\beta_{10} + \beta_{11j}x_i) \exp(-\gamma_{1j}x_i)$	θ_2	40	134,2286	0,834	0,834
IV	$(\beta_{10} + \beta_{11j}x_i) \exp(-\gamma_1x_i)$	θ_2	41	171,1564	6,928	0,008

Tabela 22 - Testes de razão de verossimilhanças para os MNGLVs encaixados da família binomial negativa tipo-2 (BN-2) com preditor de Myers ajustados aos dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Fipronil, na ausência e na presença da fração microsomal S9

Modelo	η_{1ij}	η_{2ij}	GL	$-L$	χ^2	p
I	$(\beta_{10j} + \beta_{11j}x_i) \exp(-\gamma_{1j}x_i)$	θ_{2j}	67	384,5259	–	–
II	$(\beta_{10} + \beta_{11}x_i) \exp(-\gamma_1x_i)$	θ_{2j}	70	390,9495	6,424	0,093
III	$(\beta_{10} + \beta_{11}x_i) \exp(-\gamma_1x_i)$	θ_2	71	390,9894	0,04	0,842

Tabela 23 - Testes de razão de verossimilhanças para os MNGLVs encaixados da família binomial negativa tipo-2 (BN-2) com preditor de Breslow ajustados aos dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência e na presença da fração microsomal S9

Modelo	η_{1ij}	η_{2ij}	GL	$-L$	χ^2	p
I	$e^{\beta_{10j}}(x_i + \delta)^{\beta_{11j}} e^{-\gamma_{1j}x_i}$	θ_{2j}	68	378,8221	–	–
II	$e^{\beta_{10}}(x_i + \delta)^{\beta_{11}} e^{-\gamma_1x_i}$	θ_{2j}	71	380,6637	1,842	0,606
III	$e^{\beta_{10}}(x_i + \delta)^{\beta_{11}} e^{-\gamma_1x_i}$	θ_2	72	386,4555	5,792	0,016

Tabela 24 - Testes de razão de verossimilhanças para os MLGVs encaixados da família Skellam ajustados aos dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Tiametoxam, na presença da fração microsomal S9

Modelo	η_{1i}	η_{2i}	GL	$-L$	χ^2	p
I	$\beta_{11}x_i + \beta_{12}x_i^2 + \beta_{13}x_i^3$	$\beta_{21}x_i + \beta_{22}x_i^2 + \beta_{23}x_i^3$	35	124,5228	–	–
II	$\beta_{11}x_i + \beta_{12}x_i^2 + \beta_{13}x_i^3$	$\beta_{21}x_i + \beta_{22}x_i^2$	36	124,8874	0,365	0,55
III	$\beta_{11}x_i + \beta_{12}x_i^2 + \beta_{13}x_i^3$	$\beta_{21}x_i$	37	129,8475	4,96	0,026

Tabela 25 - Testes de razão de verossimilhanças para os MLGVs encaixados da família Skellam ajustados aos dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência e na presença da fração microsomal S9

Modelo	η_{1ij}	η_{2ij}	GL	$-L$	χ^2	p
I	$\beta_{11j}x_i + \beta_{12j}x_i^2$	$\beta_{21j}x_i + \beta_{22j}x_i^2$	68	437,6433	–	–
II	$\beta_{11j}x_i + \beta_{12j}x_i^2$	$\beta_{21j}x_i + \beta_{22j}x_i^2$	70	438,2294	0,586	0,746
III	$\beta_{11}x_i + \beta_{12}x_i^2$	$\beta_{21}x_i + \beta_{22}x_i^2$	72	446,1813	7,952	0,019

Figura 47 - Resíduos e diagnósticos para o MNLGV da família binomial negativa com preditor de Myers selecionado para os dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Fipronil, na ausência e na presença da fração microsomal S9. r_i^{d*} : componentes de *deviance* estudentizados; $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 48 - Resíduos e diagnósticos para o MNLGV da família binomial negativa com preditor de Myers pré-selecionado para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Fipronil, na ausência e na presença da fração microsomal S9. $r_{(i)}^{d*}$: componentes de *deviance* estudentizados ordenados; h_{ii} : medidas de *leverage*; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta)$: medidas de curvatura local na direção da i -ésima observação

Figura 49 - Resíduos e diagnósticos para o MNLGV da família binomial negativa com preditor de Breslow pré-selecionado para os dados do ensaio de *Salmonella*/microssoma (TA100) realizado com o inseticida Tiametoxam, na ausência e na presença da fração microssomal S9. $r_{(i)}^{p*}$: resíduos de Pearson estudentizados do preditor para a média ordenados; h_{i1} : medidas de *leverage* da i -ésima observação para o preditor da média; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta_1)$: medidas de curvatura local na direção da i -ésima observação

Figura 50 - Resíduos e diagnósticos para o MLGV pré-selecionado da família Skellam para os dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Tiametoxam, na ausência e na presença da fração microsomal S9. r_i^z : resíduos padronizados; $r_{(i)}^z$: resíduos padronizados ordenados; h_{it} : medidas de *leverage* da i -ésima observação para o t -ésimo preditor; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta_1, \theta_2)$: medidas de curvatura local na direção da i -ésima observação. $\triangle$: medidas para μ_1 ; $\square$: medidas para μ_2

Figura 51 - Resíduos e diagnósticos para o MLGV pré-selecionado da família Skellam para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência e na presença da fração microssomal S9. r_i^z : resíduos padronizados; $r_{(i)}^z$: resíduos padronizados ordenados; $\hat{h}_{it}$: medidas de *leverage* da i -ésima observação para o t -ésimo preditor; $|l_{\max}|$: valores absolutos das medidas de influência local; $C_i(\theta_1, \theta_2)$: medidas de curvatura local na direção da i -ésima observação. $\triangle$: medidas para μ_1 ; $\square$: medidas para μ_2

5.3 Implicações dos modelos finais

As estimativas do modelo final selecionado para os dados da exposição da linhagem TA98 ao inseticida Fipronil estão disponíveis na Tabela 26. Os valores ajustados pelo modelo e as bandas de confiança podem ser visualizados na Figura 52, respectivamente nos painéis A e B para S9(−) e S9(+).

Nas Tabelas 27 e 28 constam as estimativas dos parâmetros dos modelos selecionados para os dados da linhagem TA98 exposta ao inseticida Tiametoxam, respectivamente para S9(−) e S9(+). Como abordado nas seções 5.1.2 e 5.2, há evidências de mecanismos de ação diferentes para a substância não-metabolizada e metabolizada, exigindo, assim, modelos distintos.

O modelo selecionado para S9(−) tem estrutura simples, de dose-resposta (ver Figura 53A) e, com relação ao parâmetro β_1 , os limites inferiores dos ICs a 95% assintótico e *bootstrap* estão bastante próximos de zero, de forma que não seria surpreendente se nenhuma ação mutagênica do Tiametoxam não-metabolizado sobre a TA98 fosse detectada em um novo experimento, principalmente se forem considerados os tamanhos amostrais modestos propostos nos protocolos.

Com relação ao modelo selecionado para S9(+), foi observada uma dinâmica complexa entre as ações de toxicidade e mutagenicidade dos produtos provenientes da metabolização do Tiametoxam (Figura 53B). O modelo ajustado prevê descréscimo inicial no número de colônias revertentes, seguido pelo aumento expressivo e um novo decaimento. Estes dois últimos processos são esperados em 5 dos 6 modelos tradicionais, então fazem-se necessárias hipóteses que tentem explicar como ocorre a diminuição no número de revertentes quando as UFCs da linhagem TA98 são expostas a baixas concentrações do Tiametoxam metabolizado via fração microssomal.

Nos estudos de Pedro (2008), foi verificado aumento da frequência de aberrações cromossômicas e de micronúcleos em doses baixas, residuais, sobre um organismo-teste eucarionte. A autora propôs que isto se deveu à incapacidade da célula de detectar a presença no meio da substância deletéria, impedindo a capacidade de impermeabilidade da membrana. Como as estruturas da membrana celular procariótica e eucariótica são distintas, é possível que o processo não esteja relacionado com a função seletiva das membranas, mas com a falha no transporte ativo para fora da célula, visto a incapacidade de detectar as moléculas mutagênicas, ou somente genotóxicas, presentes no interior dela.

Podem ser visualizadas, na Tabela 29, as estimativas dos parâmetros do modelo selecionado para os dados obtidos nos experimentos com a linhagem TA100 exposta ao inseticida Fipronil. Nessa tabela também estão disponíveis os desvios-padrões e ICs das estimativas. Os valores ajustados e as bandas de confiança obtidos por esse modelo estão disponíveis na Figura 54. Observa-se que não houve efeito da metabolização do Fipronil sobre o número de revertentes na linhagem TA100 e, ainda que baixa ($\hat{\gamma} = 0.0021$), a

toxicidade está presente tanto em S9(−) quanto em S9(+).

Finalmente, sobre os dados da exposição ao Tiametoxam da linhagem TA100, as estimativas dos parâmetros do modelo selecionado estão apresentadas na Tabela 30, incluindo os desvios-padrões e intervalos de confiança. Nesse caso, o efeito da metabolização da substância não se deu sobre o número esperado de revertentes, mas sobre a superdispersão – menor para S9(+) – de forma que os valores ajustados e as bandas de confiança, disponíveis na Figura 55, permanecem os mesmos independentemente se para S9(−) ou se para S9(+).

De acordo com o modelo selecionado para cada conjunto de dados, foram obtidas as medidas de interesse biológico (Tabela 31), descritas na seção 4.4, a dose de mutagenicidade máxima (DMM), a fração de mutagenicidade máxima (FMM), o índice de mutagenicidade máxima (IMM) e o nível de risco ambiental (NRA). Estes índices e medidas são simples e facilmente obtidos dado um modelo com bom ajuste aos dados e seu uso será pela primeira vez discutido em Butturi-Gomes et al.⁶ (em fase de elaboração).

A DMM permite avaliar o risco potencial de uma dada substância sobre sítios específicos do DNA: a DMM do Fipronil não metabolizado para a TA98 é menor que a DMM estimada desse químico para a TA100, que implica que menores quantidades do inseticida padrão são necessárias para causar mutações pontuais (*frameshift -1*), do que para causar substituição de bases. De fato, a DMM estimada do Fipronil, metabolizado ou não, para a TA100 é maior que a concentração comercializada, de forma que os riscos são bastante reduzidos para mutações por substituição de bases. Opositivamente, a DMM estimada do Tiametoxam é menor para a TA100 que para a TA98, caracterizando esse produto com maior risco de causar substituições de pares de base na molécula de DNA.

Já a FMM é uma medida que é melhor interpretada no contexto da comparação dos efeitos entre duas ou mais substâncias: quanto maior a FMM para uma dada linhagem de *S. typhimurium*, menores são os riscos contaminações residuais. Neste quesito, o Tiametoxam oferece maior risco, uma vez que apenas 30% da concentração de campo é necessária para causar a mutagenicidade máxima; contra aproximadamente 70% da concentração de campo do Fipronil. Note que a FMM pôde ser utilizada neste trabalho por terem sido avaliados químicos ambientais e em estudos com medicamentos, por exemplo, outras medidas podem e precisam ser desenvolvidas. Por outro lado, a FMM pode ser reformulada e se fazer adequada a diferentes contextos, desde que exista uma concentração de referência – como as referidas doses recomendadas para uso em campo de agroquímicos.

O IMM, uma adaptação, do ponto de vista estatístico, mais criteriosa para o IM da equação (1), é uma medida útil tanto para comparar substâncias quanto para a avaliação do potencial mutagênico: maiores valores do IMM apontam maior risco da substância em

⁶BUTTURI-GOMES, D.; FERNANDES, T.C.C.; MARIN-MORALES, M.A.; ZOCCHI, S.S. **Model-based indices and measures for the *Salmonella*/microsome assay: a case-study of the effects of Fipronil and Thiamethoxam insecticides.**

causar danos em determinadas regiões do DNA. Os maiores IMM se deram na linhagem TA98 exposta ao Fipronil, revelando, pela sobreposição dos ICs, que o potencial mutagênico, apesar das diferenças nas DMMs, é invariante ao processo de metabolização via S9, isto é, o produto padrão e os metabólitos associados ao Fipronil exercem o mesmo potencial em quantidade de dano no DNA, ainda que, para isso, sejam necessárias doses diferentes.

Por fim, a medida do NRA, que é proporcional ao IMM e inversamente proporcional à FMM, pode ser utilizada no contexto da aplicação de políticas públicas, que envolvam, particularmente, o manejo de ambientes aquáticos. O inseticida Fipronil apresentou os maiores valores de NRA, tendo seu uso não recomendado nas proximidades de corpos d'água; enquanto que o Tiametoxam apresenta maiores riscos de contaminação residual do solo que da água.

Tabela 26 - Estimativas, desvios-padrões e intervalos de confiança para os parâmetros do modelo final selecionado para os dados do ensaio de *Salmonella*/microsossoma realizado com a linhagem TA98 e exposição ao inseticida Fipronil, na ausência e na presença de um sistema de ativação metabólica (fração microsossomal S9)

Parâmetro	Estimativa	DP	IC-1		IC-2	
β_0	3,5067	0,0320	3,444	3,569	3,444	3,568
β_1	0,6141	0,0460	0,524	0,704	0,517	0,696
$\tilde{\beta}_1$	-0,2920	0,0707	-0,431	-0,153	-0,442	-0,162
γ	0,0432	0,0024	0,038	0,048	0,038	0,047
$\tilde{\gamma}$	-0,0188	0,0055	-0,030	-0,008	-0,032	-0,009
k	265,8262	46,3930	174,898	356,755	105,459	486,480

Tabela 27 - Estimativas, desvios-padrões e intervalos de confiança para os parâmetros do modelo final selecionado para os dados do ensaio de *Salmonella*/microsossoma realizado com a linhagem TA98 e exposição ao inseticida Tiametoxam, na ausência de um sistema de ativação metabólica (fração microsossomal S9)

Parâmetro	Estimativa	DP	IC-1		IC-2	
β_0	3,3308	0,0352	18,262	1,400	12,261	5,399
β_1	0,0702	0,0333	0,005	0,136	0,002	0,133

Figura 52 - Valores observados (●), ajustados (○) e bandas de confiança do modelo final selecionado para os dados do ensaio de *Salmonella*/microsossoma (TA98) realizado com o inseticida Fipronil, na ausência (A) e na presença (B) da fração microsossoma S9. Linhas tracejadas: intervalo de confiança assintótico a 95%; linhas pontilhadas: intervalo de confiança *bootstrap* a 95%

Tabela 28 - Estimativas, desvios-padrões e intervalos de confiança para os parâmetros do modelo final selecionado para os dados do ensaio de *Salmonella*/microsossoma realizado com a linhagem TA98 e exposição ao inseticida Tiametoxam, na presença de um sistema de ativação metabólica (fração microsossoma S9)

Parâmetro	Estimativa	DP	IC-1		IC-2	
β_{11}	-10,1251	4,3276	-18,607	-1,643	-12,245	-5,658
β_{12}	9,2786	4,0460	0,796	17,761	4,937	11,473
β_{13}	-1,9967	0,9620	-3,882	-0,111	-2,517	-1,045
β_{21}	-2,6621	0,1252	-2,908	-1,417	-2,688	-1,698
β_{22}	0,7742	0,0682	0,641	0,908	0,437	0,774

Tabela 29 - Estimativas, desvios-padrões e intervalos de confiança para os parâmetros do modelo final selecionado para os dados do ensaio de *Salmonella*/microsossoma realizado com a linhagem TA100 e exposição ao inseticida Fipronil, na ausência e na presença de um sistema de ativação metabólica (fração microsossoma S9)

Parâmetro	Estimativa	DP	IC-1		IC-2	
β_0	4,9932	0,0296	4,935	5,051	4,936	5,049
β_1	0,0281	0,0035	0,021	0,035	0,021	0,035
γ	0,0021	0,0003	0,002	0,003	0,001	0,003
k	21,9284	3,9303	14,225	29,632	16,369	34,503

Figura 53 - Valores observados ($\bullet$), ajustados ($\circ$) e bandas de confiança dos modelos finais selecionados para os dados do ensaio de *Salmonella*/microsoma (TA98) realizado com o inseticida Tiametoxam, na ausência (A) e na presença (B) da fração microsomal S9. Linhas tracejadas: intervalo de confiança assintótico a 95%; linhas pontilhadas: intervalo de confiança *bootstrap* a 95%

Figura 54 - Valores observados ($\bullet$), ajustados ($\circ$) e bandas de confiança do modelo final selecionado para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Fipronil, na ausência e na presença da fração microsomal S9. Linhas tracejadas: intervalo de confiança assintótico a 95%; linhas pontilhadas: intervalo de confiança *bootstrap* a 95%

Tabela 30 - Estimativas, desvios-padrões e intervalos de confiança para os parâmetros do modelo final selecionado para os dados do ensaio de *Salmonella*/microsoma realizado com a linhagem TA100 e exposição ao inseticida Tiametoxam, na ausência e na presença de um sistema de ativação metabólica (fração microsomal S9)

Parâmetro	Estimativa	DP	IC-1		IC-2	
β_{10}	4,8304	0,0389	4,754	4,907	4,753	4,906
β_{11}	2,0249	0,4015	1,238	2,812	1,228	2,854
γ_1	1,0423	0,2008	0,649	1,436	0,659	1,463
θ_{21}	1,9459	0,2461	1,464	2,428	1,536	2,523
θ_{22}	3,2060	0,2517	2,713	3,699	2,802	3,969

Figura 55 - Valores observados ($\bullet$), ajustados ($\circ$) e bandas de confiança do modelo final selecionado para os dados do ensaio de *Salmonella*/microsoma (TA100) realizado com o inseticida Tiametoxam, na ausência e na presença da fração microsomal S9. Linhas tracejadas: intervalo de confiança assintótico a 95%; linhas pontilhadas: intervalo de confiança *bootstrap* a 95%

Tabela 31 - Estimativas de máxima verossimilhança das medidas de interesse biológico estimadas pelo modelo selecionado, de acordo com a linhagem de *S. typhimurium*, inseticida e sistema de ativação metabólica avaliados em cada experimento. SAM S9: fração microsomal ausente (–) ou presente (+). DMM: dose de mutagenicidade máxima em g/L. FMM: fração de mutagenicidade máxima. IMM: índice de mutagenicidade máxima. NRA: nível de risco ambiental. Os valores entre colchetes, $[\cdot ; \cdot]$, denotam os limites inferior e superior do intervalo de confiança *bootstrap* a 95%

Linhagem	Inseticida	SAM	DMM	FMM	IMM	NRA
TA98	Fipronil	S9(–)	17,45073 [16,1; 19,7]	0,7 [0,64; 0,79]	24,26 [19,9; 29,9]	34,76 [25,6; 46,2]
TA98	Fipronil	S9(+)	30,14585 [22,26; 73,18]	1,2 [0,89; 2,93]	17,01 [14,92; 73,89]	14,1 [12,46; 25,1]
TA98	Tiametoxam	S9(–)	–	–	–	–
TA98	Tiametoxam	S9(+)	2,385728 [2,31; 2,4]	0,79 [0,77; 0,80]	4,53 [1,07; 8,83]	5,7 [1,34; 11,10]
TA100	Fipronil	S9(–)	294,908 [230,6; 447,1]	11,79 [9,22; 17,88]	8,39 [6,76; 13,46]	0,71 [0,61; 0,86]
TA100	Fipronil	S9(+)	294,908 [230,6; 447,1]	11,79 [9,22; 17,88]	8,39 [6,76; 13,46]	0,71 [0,61; 0,86]
TA100	Tiametoxam	S9(–)	0,942699 [0,77; 1,09]	0,31 [0,26; 0,36]	1,44 [1,2; 1,72]	4,57 [3,94; 5,27]
TA100	Tiametoxam	S9(+)	0,942699 [0,77; 1,09]	0,31 [0,26; 0,36]	1,44 [1,2; 1,72]	4,57 [3,94; 5,27]

6 CONCLUSÕES

O pacote “ames”, desenvolvido em linguagem R no presente trabalho, é uma ferramenta suficientemente completa para o ajuste, diagnóstico e seleção de modelos para os do ensaio de *Salmonella*/microsoma. Certamente, não se trata de uma plataforma ampla, como as disponíveis nas funções `glm()` e `vglm()`, e não pretende ser, pois trata de casos particulares, para os quais evidências empíricas já tenham sido obtidas. Assim, é de importância estratégica, visto que há carência tanto de *softwares* gratuitos, de código aberto e de fácil acesso para realizar a análise de dados desse ensaio quanto da própria aplicação de métodos estatísticos na literatura. As perspectivas futuras, como a manutenção, atualização e otimização dos procedimentos no pacote “ames” ficarão a cargo dos autores, bem como a implementação de possíveis outras ferramentas para ajuste de modelos, diagnóstico, seleção e inferência.

A introdução da classe dos MNLGVs se apresentou relevante para os dados do TA, visto que os melhores ajustes foram obtidos ou por casos particulares (um MLGV para os dados da linhagem TA98 exposta ao inseticida Tiametoxam na presença de S9, e um MNLG para os dados da linhagem TA100 exposta ao inseticida Fipronil) ou por MNLGVs estritos (dados da linhagem TA100 exposta ao inseticida Tiametoxam). Essa nova classe de modelos merece atenção em estudos, tanto teóricos quanto práticos, pois, devido a sua formulação ampla, novas aplicações, além do próprio TA, poderão ser encontradas.

Apesar das dificuldades numéricas e das limitações computacionais ao se trabalhar com a distribuição de Skellam, os modelos com esse componente aleatório se mostraram relevantes para os dados obtidos neste trabalho e têm um potencial para apresentar bons ajustes a novos conjuntos de dados provenientes do teste de *Salmonella*/microsoma. Além disso, os MLGVs da família Skellam apresentam características tanto de modelos empíricos quanto de modelos mecanicistas, enfatizando sua importância na área.

Foram selecionados diferentes modelos, variando ou com relação à linhagem da bactéria, ou com a substância sob investigação, ou com o nível do fator S9, ou, ainda, com uma combinação desses quesitos. Como cada modelo descreve os comportamentos tóxicos e mutagênicos de forma diferente, a comparação dos riscos do uso de duas ou mais substâncias diferentes deve ser pautada nas medidas FMM e NRA, e, para avaliar o mecanismo e o potencial mutagênicos de ação de uma dada substância, as medidas DMM e IMM são recomendadas sobre a medida clássica, IM. Levando isto em consideração, pode-se concluir que o Tiametoxam tem maior potencial mutagênico para a linhagem TA100 (DMM e FMM baixas) e que o uso do Fipronil é, comparativamente, mais arriscado (NRA e IMM altos).

Por fim, é necessário ressaltar que, para verificar os potenciais genotóxico e mutagênico dos inseticidas investigados no presente trabalho, mais tipos de ensaios biológicos devem ser aplicados utilizando outros organismos-teste, como os testes do micrônúcleo

e de aberrações cromossômicas e o ensaio do cometa. Além disso, recomenda-se o uso de maiores tamanhos amostrais do que aqueles propostos nos protocolos para o TA, de forma que as conclusões obtidas a partir dos dados e dos modelos ajustados neste trabalho poderão ser melhoradas, principalmente no que se refere a formas mais confiáveis para a variância e maior precisão para as distribuições preditivas.

REFERÊNCIAS

ABRAMOWITZ, M.; STEGUN, I. A. (Ed.). **Handbook of mathematical functions**. New York: Dover Publications, 1972. 1024 p.

AITKIN, M. NPML estimation of the mixing distribution in general statistical models with unobserved random effects. In: SEEGER, G. U. H.; FRANCIS, B. J.; HATZINGER, R.; STECKEL-BERGER, G. (Ed.). **Statistical modelling**. New York: Springer, 1995. p. 1–9.

AKAIKE, H. Information theory as an extension of the maximum likelihood principle. In: PETROV, B. N.; CSAKI, F. (Ed.). **Second International Symposium on Information Theory**. Budapest: Akademiai Kiado, 1973. p. 267–281.

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: KOTZ, S.; JOHNSON, N. L. (Ed.). **Breakthroughs in Statistics: Foundations and Basic Theory**. New York: Springer, 1992. p. 610–624.

ALZAID, A. A.; OMAIR, M. A. On the Poisson difference distribution inference and applications. **Bulletin of the Malaysian Mathematical Sciences Society**, Penang, v. 33, n. 1, p. 17–45, 2010.

AMES, B. N.; DURSTON, W. E.; YAMASAKI, E.; LEE, F. D. Carcinogens are mutagens: a simple test system combining liver homogenates for activation and bacteria for detection. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 70, n. 8, p. 2281–2285, 1973.

AMES, B. N.; LEE, F. D.; DURSTON, W. E. An improved bacterial test system for the detection and classification of mutagens and carcinogens. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 70, n. 3, p. 782–786, 1973.

BAKER, R. J.; NELDER, J. A. **The GLIM System: Release 3**. Oxford, 1978.

BATES, D. M.; MAECHLER, M. **Matrix**: Sparse and dense matrix classes and methods. [S.l.], 2015. R package version 1.2-0. Disponível em: <<http://CRAN.R-project.org/package=Matrix>>. Acesso em: 05 maio 2015.

BATES, D. M.; WATTS, D. G. Relative curvature measures of nonlinearity. **Journal of the Royal Statistical Society: Series B**, Oxford, v. 42, n. 1, p. 1–25, 1980.

BATES, D. M.; WATTS, D. G. **Nonlinear regression analysis and its applications**. New York: Wiley, 1988.

BERNSTEIN, L.; KALDOR, J.; MacCANN, J.; PIKE, M. C. An empirical approach to the statistical analysis of mutagenesis data from the *Salmonella* test. **Mutation Research**, Toronto, v. 97, p. 267–281, 1982.

BRESLOW, N. Tests of hypotheses in overdispersed Poisson regression and other quasilielihood models. **Journal of the American Statistical Association**, Alexandria, v. 85, n. 410, p. 565–571, 1990.

BRESLOW, N. E. Extra-poisson variation in log-linear models. **Journal of the Royal Statistical Society: Series C**, Oxford, v. 33, n. 1, p. 38–44, 1984.

BURNHAM, K. P.; ANDERSEN, D. R. **Model selection and multimodel inference**. 2nd. ed. New York: Springer, 2002. 488 p.

BURNHAM, K. P.; ANDERSEN, D. R. Multimodel inference: understanding AIC and BIC in model selection. **Sociological Methods & Research**, Thousand Oaks, v. 33, n. 2, p. 261–304, 2004.

CASTILLO, J. del; PÉREZ-CASANY, M. Overdispersed and underdispersed Poisson generalizations. **Journal of Statistical Planning and Inference**, Singapore, v. 134, p. 486–500, 2005.

CAUTRES, B.; HEATH, A. F.; FIRTH, D. Class, religion and vote in Britain and France. **La Lettre de la Maison Francaise**, v. 8, p. 30–45, 1998.

CAVALCANTI, A. B.; CORDEIRO, G. M.; BOTTER, D. A.; BARROSO, L. P. Asymptotic skewness in exponential family nonlinear models. **Communications in Statistics – Theory and Methods**, Alexandria, v. 38, n. 14, p. 2285–2287, 2009.

CLAXTON, L.; CREASON, J.; NADER, J. A.; POTEAT, W.; ORR, J. D. GENETOX manager for bacterial mutagenicity assays: a personal computer and mini computer system. **Mutation Research**, Toronto, v. 342, p. 87–94, 1995.

CONSUL, P. C.; JAIN, G. C. A generalization of the Poisson distribution. **Technometrics**, Alexandria, v. 15, n. 4, p. 791–799, 1973.

COOK, R. D. Assessment of local influence. **Journal of the Royal Statistical Society: Series B**, Oxford, v. 48, n. 2, p. 133–169, 1986.

CORDEIRO, G. M.; DEMÉTRIO, C. G. B. Corrected estimators in extended quasi-likelihood models. **Communications in Statistics – Theory and Methods**, Alexandria, v. 37, p. 873–880, 2008.

CORDEIRO, G. M.; McCULLAGH, P. Bias correction in generalized linear models. **Journal of the Royal Statistical Society: Series B**, Oxford, v. 53, n. 3, p. 629–643, 1991.

CORDEIRO, G. M.; PAULA, G. A. Improved likelihood ratio statistics for exponential family nonlinear models. **Biometrika**, Oxford, v. 76, n. 1, p. 93–100, 1989.

CORDEIRO, G. M.; PREVIDELLI, I.; SAMOHYL, R. W. Bias corrected maximum likelihood estimators in nonlinear overdispersed models. **Brazilian Journal of Probability and Statistics**, v. 22, n. 2, p. 105–118, 2008.

CORDEIRO, G. M.; SANTANA, R. G. Covariance matrix formula for exponential family nonlinear models. **Communications in Statistics – Theory and Methods**, Alexandria, v. 37, n. 17, p. 2724–2734, 2008.

CORDEIRO, G. M.; UDO, M. C. T. Bias correction in generalized nonlinear models with dispersion covariates. **Communications in Statistics – Theory and Methods**, Alexandria, v. 37, n. 14, p. 2219–2225, 2008.

DARBY, S. C.; ELLIS, M. T. A test for synergism between two drugs. **Journal of the Royal Statistical Society: Series C**, Oxford, v. 25, n. 3, p. 296–299, 1976.

DAVIDIAN, M.; CARROLL, R. J. Variance function estimation. **Journal of the American Statistical Association**, Alexandria, v. 82, n. 400, p. 1079–1091, 1987.

DAVIDIAN, M.; GILTINAN, D. M. **Nonlinear models for repeated measurement data**. New York: Chapman and Hall/CRC, 1995.

DEAN, C.; LAWLESS, J. F.; WILLMOT, G. E. A mixed Poisson-Inverse-Gaussian regression model. **The Canadian Journal of Statistics**, Ottawa, v. 17, n. 2, p. 171–181, 1989.

DEAN, C. B. Testing for overdispersion in Poisson and binomial regression models. **Journal of the American Statistical Society**, Alexandria, v. 87, n. 418, p. 451–457, 1992.

DEAN, C. B.; LAWLESS, J. F. Tests for detecting overdispersion in Poisson regression models. **Journal of the American Statistical Association**, Alexandria, v. 84, n. 406, p. 467–472, 1989.

DEARFIELD, K. L.; CIMINO, M. C.; MCCARROLL, N. E.; MAUERC, I.; VALCOVIC, L. R. Genotoxicity risk assessment: a proposed classification strategy. **Mutation Research**, Toronto, v. 521, n. 1–2, p. 121–135, 2002.

deLEEUW, J. Introduction to Akaike (1973) information theory and extension of the maximum likelihood principle. In: KOTZ, S.; JOHNSON, N. L. (Ed.). **Breakthroughs in Statistics: Foundations and Basic Theory**. New York: Springer, 1992. p. 599–609.

DUNN, P. **tweedie**: Tweedie exponential family models. [S.l.], 2014. R package version 2.2.1. Disponível em: <<http://CRAN.R-project.org/package=tweedie>>. Acesso em: 12 nov. 2014.

ECKARDT, F.; HAYNES, R. H. Kinetics of mutation induction by ultraviolet in excision-deficient yeast. **Genetics**, v. 85, p. 225–247, 1977.

EFRON, B.; TIBSHIRANI, R. J. **An introduction to the bootstrap**. Boca Raton: Chapman & Hall, 1993. 436 p.

GALLANT, A. R. Nonlinear regression. **The American Statistician**, Alexandria, v. 29, n. 2, p. 73–81, 1975.

GANIO, L. M.; SCHAFER, D. W. Diagnostics for overdispersion. **Journal of the American Statistical Association**, Alexandria, v. 87, n. 419, p. 795–804, 1992.

GIBBONS, D.; MORRISSEY, C.; MINEAU, P. A review of the direct and indirect effects of neonicotinoids and fipronil on vertebrate wildlife. **Environmental Science and Pollution Research**, Heidelberg, v. 22, p. 103–118, 2015.

GOULET, V.; DUTANG, C.; MAECHLER, M.; FIRTH, D.; SHAPIRA, M.; STADELMANN, M. **expm**: Matrix exponential. [S.l.], 2014. R package version 0.99-1.1. Disponível em: <<http://CRAN.R-project.org/package=expm>>. Acesso em: 05 maio 2015.

GREEN, P. J. Iteratively Reweighted Least Squares for maximum likelihood estimation, and some robust and resistant alternatives. **Journal of the Royal Statistical Society: Series B**, Oxford, v. 46, n. 2, p. 149–192, 1984.

GREENE, W. Functional forms for the negative binomial model for count data. **Economics Letters**, Philadelphia, v. 99, p. 585–590, 2008.

HAMADA, C.; WADA, T.; SAKAMOTO, Y. Statistical characterization of negative control data in the ames *Salmonella*/microsome test. **Environmental Health Perspectives**, v. 102, p. 115–119, 1994.

HANKIN, R. K. S. **hypergeo**: The hypergeometric function. [S.l.], 2015. R-package version 1.2-9. Disponível em: <<http://CRAN.R-project.org/package=hypergeo>>. Acesso em: 16 junho 2015.

HANKIN, R. K. S.; LEE, A. A new family of non-negative distributions. **Australia & New Zealand Journal of Statistics**, Sydney, v. 48, n. 1, p. 67–78, 2006.

HINDE, J.; DEMÉTRIO, C. G. B. Overdispersion: models and estimation. **Computational Statistics & Data Analysis**, The Hague, v. 27, p. 151–170, 1998.

HOUGAARD, P.; LEE, M.-L. T.; WHITMORE, G. A. Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes. **Biometrics**, Washington, v. 53, n. 4, p. 1225–1238, 1997.

HURVICH, C. M.; TSAI, C.-L. Regression and time series model selection in small samples. **Biometrika**, Oxford, v. 76, n. 2, p. 297–307, 1989.

HURVICH, C. M.; TSAI, C.-L. Bias of the corrected AIC criterion for underfitted regression and time series models. **Biometrika**, Oxford, v. 78, n. 3, p. 499–509, 1991.

HURVICH, C. M.; TSAI, C.-L. Model selection for extended quasi-likelihood models in small samples. **Biometrics**, Washington, v. 51, n. 3, p. 1077–1084, 1995.

IRWIN, J. O. The frequency distribution of the difference between two independent variates following the same Poisson distribution. **Journal of the Royal Statistical Society: Series A**, Oxford, v. 100, n. 3, p. 415–416, 1937.

JEONG, E.; FREEMAN, H. S.; CLAXTON, L. D. Synthesis and characterization of selected 4,4'-diaminoalkoxyazobenzenes. **Dyes and Pigments**, Shanghai, v. 87, n. 2, p. 100–108, 2010.

JOE, H.; ZHU, R. Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. **Biometrical Journal**, Hannover, v. 47, n. 2, p. 219–229, 2005.

JØRGENSEN, B. Maximum likelihood estimation and large-sample inference for generalized and nonlinear regression models. **Biometrika**, Oxford, v. 70, n. 1, p. 19–28, 1983.

JØRGENSEN, B. **The theory of dispersion models**. London: Chapman & Hall, 1997. 237 p.

- KIM, B. S.; MARGOLIN, B. H. Statistical methods of the Ames *Salmonella* assay: a review. **Mutation Research**, Toronto, v. 436, p. 113–122, 1999.
- KOKONENDJI, C. C.; DEMÉTRIO, C. G. B.; ZOCCHI, S. S. On Hinde-Demétrio regression models for overdispersed count data. **Statistical Methodology**, Cambridge, v. 4, p. 277–291, 2007.
- KOKONENDJI, C. C.; DOSSOU-GBÉTÉ, S.; DEMÉTRIO, C. G. B. Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes. **Statistics and Operations Research Transactions**, Barcelona, v. 28, n. 2, p. 201–214, 2004.
- KREWSKI, D.; FRANKLIN, C. (Ed.). **Statistics in toxicology**. New York: Gordon and Breach, 1991. 677 p.
- KREWSKI, D.; LEROUX, B. G.; BLEUR, S. R.; BROEKHAVEN, L. H. Modelling the Ames *Salmonella*/microsome assay. **Biometrics**, Washington, v. 49, p. 499–510, 1993.
- KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. **Annals of Mathematical Statistics**, Beachwood, v. 22, n. 2, p. 79–86, 1951.
- LAMBERT, D.; ROEDER, K. Overdispersion diagnostics for generalized linear models. **Journal of the American Statistical Association**, Alexandria, v. 90, n. 432, p. 1225–1236, 1995.
- LAWLESS, J. F. Negative binomial and mixed Poisson regression. **The Canadian Journal of Statistics**, Ottawa, v. 15, n. 3, p. 209–225, 1987.
- LEROUX, B. G.; KREWSKI, D. AMESFIT: a microcomputer program for fitting linear-exponential dose-response models in the Ames *Salmonella* assay. **Environmental and Molecular Mutagenesis**, Malden, v. 22, p. 78–84, 1993.
- LESAFFRE, E.; VERBEKE, G. Local influence in linear mixed models. **Biometrics**, Washington, v. 54, n. 2, p. 570–582, 1998.
- LI, B. On profile likelihood: comment. **Journal of the American Statistical Association**, Alexandria, v. 95, n. 450, p. 472–474, 2000.
- LIN, J.-G.; LI, Y. Testing for departures from nominal dispersion in generalized nonlinear models with varying dispersion and/or additive random effects. **Journal of Statistical Computation and Simulation**, Alexandria, v. 78, n. 10, p. 925–939, 2008.
- LIN, L.; ZHANG, R. Profile quasi-likelihood. **Statistics & Probability Letters**, Cambridge, v. 56, p. 147–154, 2002.
- MACASKILL, G. T. A note on adjusted profile likelihoods in non-linear regression. **Journal of the Royal Statistical Society: Series B**, Oxford, v. 55, n. 1, p. 125–131, 1993.
- MAECHLER, M. **Bessel**: Bessel functions computations and approximations. [S.l.], 2015. R-package version 0.6-0/r24. Disponível em: <<http://R-Forge.R-project.org/projects/specfun/>>. Acesso em: 05 maio 2015.

MARGOLIN, B. H.; KAPLAN, N.; ZEIGER, E. Statistical analysis of the Ames *Salmonella*/microsome test. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 78, n. 6, p. 3779–3783, 1981.

McCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. 2nd. ed. London: Chapman and Hall, 1989. 511 p.

McCULLAGH, P.; TIBSHIRANI, R. A simple method for the adjustment of profile likelihoods. **Journal of the Royal Statistical Society: Series B**, Oxford, v. 52, n. 2, p. 325–344, 1993.

MOLLA, D. T.; MUNISWAMY, B. Power of tests for overdispersion parameter in negative binomial regression model. **International Organization of Scientific Research Journal of Mathematics**, New Delhi, v. 1, n. 4, p. 29–36, 2012.

MOOLGAVKAR, S. H.; LUSTBADER, E. D.; VENZON, D. J. Geometric approach to nonlinear regression diagnostics with application to matched case-control studies. **Annals of Statistics**, Philadelphia, v. 12, n. 3, p. 816–826, 1984.

MORTELMANS, K.; ZEIGER, E. The Ames *Salmonella*/microsome mutagenicity assay. **Mutation Research**, Toronto, v. 455, p. 29–60, 2000.

MURPHY, S. A.; VAART, A. W. van der. A simple method for the adjustment of profile likelihoods. **Journal of the Royal Statistical Society: Series B**, Oxford, v. 52, n. 2, p. 325–344, 1993.

MYERS, L. E.; SAXTON, N. H.; SOUTHERLAND, L. I.; WOLFF, T. J. Regression analysis of Ames test data. **Environmental and Molecular Mutagenesis**, Malden, v. 3, p. 575–586, 1981.

NAKASHIMA, E. Some methods for estimation in negative binomial model. **Annals of the Institute of Statistical Mathematics**, Tokyo, v. 49, n. 1, p. 101–115, 1997.

NELDER, J. A.; LEE, Y. Likelihood, quasi-likelihood and pseudolikelihood: some comparisons. **Journal of the Royal Statistical Society: Series B**, Oxford, v. 54, n. 1, p. 273–284, 1992.

NELDER, J. A.; PREGIBON, D. An extended quasi-likelihood function. **Biometrika**, Oxford, v. 74, n. 2, p. 221–232, 1987.

PAULA, G. A. Bias correction for exponential family nonlinear models. **Journal of Statistical Computation and Simulation**, Alexandria, v. 40, n. 1–2, p. 43–54, 1992.

PEDRO, J. **Deteção da citotoxicidade, genotoxicidade e mutagenicidade do inseticida Fipronil no organismo teste *Allium cepa***. 103 p. Dissertação (Mestrado em Biologia Celular e Molecular) – Instituto de Biociências, Universidade Estadual Paulista, Rio Claro, 2008. Disponível em: <<http://hdl.handle.net/11449/87704>>.

PEREIRA, M. **Tiametoxam em plantas de cana-de-açúcar, feijoeiro, soja, laranjeira e cafeeiro: parâmetros de desenvolvimento e aspectos bioquímicos**. 124 p. Tese (Doutorado em Fitotecnia) – Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2010.

- PRENTICE, R. L. Correlated binary regression with covariates specific to each binary observation. **Biometrics**, Washington, v. 44, p. 1033–1048, 1988.
- PRENTICE, R. L.; ZHAO, L. P. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. **Biometrics**, Washington, v. 47, n. 3, p. 825–839, 1991.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna: R Foundation for Statistical Computing, 2015. Disponível em: <<http://www.R-project.org/>>. Acesso em: 20 de maio de 2015.
- RATKOWSKY, D. A. **Nonlinear regression modelling – a unified practical approach**. New York: Marcel Dekker Inc., 1983.
- RATKOWSKY, D. A. **Handbook of nonlinear regression models**. New York: Marcel Dekker Inc., 1989.
- ROSA, A. S. **Efeitos da exposição de *Bombus terrestris audax*, *Apis mellifera carnica* e *Scaptotrigona bipunctata* ao neonicotinóide tiametoxam e uso de *Scaptotrigona aff. depilis* como bioindicador**. 143 p. Tese (Doutorado em Entomologia) – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2014.
- SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, Philadelphia, v. 6, p. 461–464, 1978.
- SHANNON, C. E. A mathematical theory of communication. **Bell System Technical Journal**, Bedminster, v. 27, p. 279–423, 1948.
- SHIBATA, R. Bootstrap estimate of Kullback-Leibler information for model selection. **Statistica Sinica**, Statesboro, v. 7, p. 375–394, 1997.
- SIMAS, A. B.; CORDEIRO, G. M. Adjusted Pearson residuals in exponential family nonlinear models. **Journal of Statistical Computation and Simulation**, Alexandria, v. 79, n. 4, p. 411–425, 2009.
- SIMPSON, D. G.; MARGOLIN, B. H. Recursive nonparametric testing for dose–response relationships subject to downturns at high doses. **Biometrika**, Oxford, v. 73, p. 589–596, 1986.
- SIMPSON, D. G.; MARGOLIN, B. H. Nonparametric testing for dose–response curves subject to downturns: Asymptotic power considerations. **Annals of Statistics**, Philadelphia, v. 18, p. 373–590, 1990.
- SKELLAM, J. G. The frequency distribution of the difference between two Poisson variates belonging to different populations. **Journal of the Royal Statistical Society: Series B**, Oxford, v. 109, n. 3, p. 296, 1946.
- SLUIJS, J. P. van der; AMARAL-ROGERS, V.; BELZUNCES, L. P.; LEXMOND, M. F. I. J. B. van; BONMATIN, J.-M.; CHAGNON, M.; DOWNS, C. A.; FURLAN, L.; GIBBONS, D. W.; GIORIO, C.; GIROLAMI, V.; GOULSON, D.; KREUTZWEISER, D. P.; KRUPKE, C.; LIESS, M.; LONG, E.; MCFIELD, M.; MINEAU, P.; MITCHELL,

E. A. D.; MORRISSEY, C. A.; NOOME, D. A.; PISA, L.; SETTELE, J.; SIMONDELSON, N.; STARK, J. D.; TAPPARO, A.; DYCK, H. V.; PRAAGH, J. van; WHITEHORN, P. R.; WIEMERS, M. Conclusions of the Worldwide Integrated Assessment on the risks of neonicotinoids and fipronil to biodiversity and ecosystem functioning. **Environmental Science and Pollution Research**, Heidelberg, v. 22, p. 148–154, 2015.

SMITH, P. J.; HEITJAN, D. F. Testing and adjusting for departures from nominal dispersion in generalized linear models. **Journal of the Royal Statistical Society: Series C**, Oxford, v. 42, n. 1, p. 31–41, 1993.

SNEDECOR, G. W.; COCHRAN, W. G. **Statistical Methods**. Ames: Iowa State University Press, 1980.

SOETAERT, K. **rootSolve**: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations. [S.l.], 2009. R-package version 1.6. Disponível em: <<http://CRAN.R-project.org/package=rootSolve>>. Acesso em: 6 fev. 2014.

SOETAERT, K.; HERMAN, P. M. J. **A Practical Guide to Ecological Modelling: Using r as a simulation platform**. New York: Springer, 2009. 372 p.

STEAD, A. G.; HASSELBLAD, V.; CREASON, J. P.; CLAXTON, L. Modelling the Ames test. **Mutation Research**, Toronto, v. 85, p. 13–27, 1981.

SUEIRO, R. A.; GARRIDO, M. J.; ARAUJO, M. Mutagenic assessment of prestige fuel oil spilled on the shore and submitted to field trials of bioremediation. **Science of the Total Environment**, Cambridge, v. 409, n. 23, p. 4973–4978, 2011.

SVETLIZA, C. F.; PAULA, G. A. Diagnostics in nonlinear negative binomial models. **Communications in Statistics – Theory and Methods**, Alexandria, v. 32, n. 6, p. 1227–1250, 2003.

TAFURT, Y.; MARIN, M. A. Principales mecanismos de reparación de daños en la molécula de ADN. **Revista Biosalud**, Manizales, v. 13, n. 2, 2014.

TAKEUCHI, K. Distribution of information statistics and a criterion of model fitting. **Suri-Kagaku (Mathematical Sciences)**, Sakai, v. 153, p. 12–18, 1976.

TSALLIS, C. Possible generalization of Boltzmann-Gibbs statistics. **Journal of Statistical Physics**, Piscataway, v. 52, p. 479–487, 1988.

TURNER, H. L.; FIRTH, D. Tools for computing generalized nonlinear models in R. **Statistical Computing and Graphics**, Alexandria, v. 18, n. 1, p. 11–16, 2007.

UMBUZEIRO, G. A.; KUMMROW, D.; MORALES, D. A.; ALVES, D. K. M.; LIM, H.; JARVIS, I. W. H.; BERGVALL, C.; WESTERHOLM, R.; STENIUS, U.; DREIJ, K. Sensitivity of salmonella yg5161 for detecting pah-associated mutagenicity in air particulate matter. **Environmental and Molecular Mutagenesis**, Malden, v. 55, n. 6, p. 510–517, 2014.

UMBUZEIRO, G. A.; VARGAS, V. M. F.; FELZENSZWALB, I.; HENRIQUES, J. A. P.; VARANDA, E. Teste de mutação reversa com *Salmonella typhimurium*

- (**Teste de Ames, Ensaio *Salmonella*/microsoma**). Ribeirão Preto: Sociedade Brasileira de Mutagenese, Carcinogenese e Teratogenese Ambiental (SBMCTA), 2004. Disponível em: <<http://www.sbmcta.org.br/documentos/>>. Acesso em: 10 jan. 2012.
- VENABLES, W. N.; RIPLEY, B. D. **Modern applied statistics with S**. New York: Springer, 2002. 495 p.
- VONESH, E. F. **Generalized linear and nonlinear models for correlated data**. Cary: SAS Institute Inc., 2012.
- WAHRENDORF, J.; MAHON, G. A. T.; SCHUMACHER, M. A nonparametric approach to the statistical analysis of mutagenicity data. **Mutation Research**, Toronto, v. 147, p. 5–13, 1985.
- WARREN, S. H.; CLAXTON, L. D.; DILIBERTO, J.; HUGHES, T. J.; SWANK, A.; KUSNIERZ, D. H.; MARSHALL, V.; DeMARINI, D. M. Survey of the mutagenicity of surface water, sediments, and drinking water from the penobscot indian nation. **Chemosphere**, Amsterdam, v. 120, p. 690–696, 2015.
- WEDDERBURN, R. W. M. Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. **Biometrika**, Oxford, v. 61, n. 3, p. 439–447, 1974.
- WEI, B.-C. **Exponential family nonlinear models**. Singapore: Springer, 1998. 230 p.
- WEINSTEIN, D.; LEWINSON, T. M. A statistical treatment of the Ames mutagenicity assay. **Mutation Research**, Toronto, v. 51, p. 433–434, 1978.
- WELCH, L. F.; ADAMS, W. E.; CORMAN, J. L. Yield response surfaces, isoquants and economic fertilizer optima for coastal Bermudagrass. **Agronomy Journal**, Madison, v. 55, p. 63–67, 1963.
- WICKHAM, H.; CHANG, W. **devtools: Tools to make developing R packages easier**. [S.l.], 2015. R package version 1.8.0. Disponível em: <<http://CRAN.R-project.org/package=devtools>>. Acesso em: 31 agosto 2015.
- WUERTZ, D. **fAsianOptions: EBM and Asian Option Valuation**. [S.l.], 2013. R-package version 3010.79. Disponível em: <<http://CRAN.R-project.org/package=fAsianOptions>>. Acesso em: 16 junho 2015.
- XIANG, L.; LEE, A. H. Sensitivity of test for overdispersion in Poisson regression. **Biometrical Journal**, Hannover, v. 47, n. 2, p. 167–176, 2005.
- YANG, Z.; HARDIN, J. W.; ADDY, C. L. A note on Dean's overdispersion test. **Journal Statistical Planning and Inference**, Singapore, v. 139, p. 3675–3678, 2009.
- YANG, Z.; HARDIN, J. W.; ADDY, C. L. A score test for overdispersion in Poisson regression based on the generalized Poisson-2 model. **Journal Statistical Planning and Inference**, Singapore, v. 139, p. 1514–1521, 2009.
- YANG, Z.; HARDIN, J. W.; ADDY, C. L.; VUONG, Q. H. Testing approaches for overdispersion in Poisson regression versus the generalized Poisson model. **Biometrical Journal**, Hannover, v. 49, n. 4, p. 565–584, 2007.

- YEE, T. W. The VGAM package for categorical data analysis. **Journal of Statistical Software**, Alexandria, v. 32, n. 10, p. 1–34, 2010. Disponível em: <http://www.jstatsoft.org/v32/i10/>. Acesso em: 8 jan. 2014.
- YEE, T. W. **VGAM**: Vector Generalized Linear and Additive Models. [S.l.], 2014. R package version 0.9-6. Disponível em: <http://CRAN.R-project.org/package=VGAM>. Acesso em: 6 fev. 2014.
- YEE, T. W.; HASTIE, T. J. Reduced-rank vector generalized linear models. **Statistical Modelling**, Thousand Oaks, v. 3, p. 15–41, 2003.
- YEE, T. W.; WILD, C. J. Vector generalized additive models. **Journal of the Royal Statistical Society: Series B**, Oxford, v. 58, n. 3, p. 481–493, 1996.
- ZAMANI, H.; ISMAIL, N. Functional form for the generalized Poisson regression model. **Communications in Statistics – Theory and Methods**, Alexandria, v. 41, n. 20, p. 3666–3675, 2012.
- ZAMANI, H.; ISMAIL, N. Score test for testing Poisson regression against generalized Poisson alternatives. **Proceeding of the 20th National Symposium on Mathematical Sciences**, Putrajaya, v. 1522, p. 1205–1212, 2013.

APÊNDICES

Apêndice A: Estudo dos preditores tradicionais

O preditor do modelo de Breslow (equação (9)) é contínuo em todo o domínio de $\{f(x, \delta, \boldsymbol{\theta}) : \mathbb{R}_+ \times \mathbb{R}_+^* \times \mathbb{R}_+^3\}$. A primeira derivada em relação a x é

$$\frac{\partial f}{\partial x} = \frac{\beta_1 e^{\beta_0 - \gamma x} (x + \delta)^{\beta_1}}{x + \delta} - \gamma e^{\beta_0 - \gamma x} (x + \delta)^{\beta_1},$$

cuja solução para x é

$$\frac{\partial f}{\partial x} = 0 \Leftrightarrow x^* = \frac{\beta_1 - \gamma \delta}{\gamma}.$$

Pode-se verificar que $f(x^*, \delta, \boldsymbol{\theta})$ é um ponto de máximo, mas quando $\gamma \gg \beta_1 \Rightarrow x^* \notin \mathbb{R}_+$. Mais detalhes sobre os formatos adquiridos pelo preditor para diferentes valores dos parâmetros estão disponíveis na Figura 56A.

Quanto aos limites do preditor, existe uma desvantagem: uma vez que o valor do intercepto do modelo deve refletir apenas o número de revertentes espontâneos (β_0), a quantidade δ^{β_1} , particularmente β_1 , o potencial mutagênico da substância, deveria ser nula no controle negativo.

$$\lim_{x \rightarrow \infty} f(x, \boldsymbol{\theta}) = 0 \quad \text{e} \quad \lim_{x \rightarrow 0^+} f(x, \boldsymbol{\theta}) = e^{\beta_0} \delta^{\beta_1}$$

O preditor do modelo de Krewski (equação (10)) não é contínuo em todo o domínio $\{f(x, \alpha, \boldsymbol{\theta}) : x \geq 0, \alpha > 1, \boldsymbol{\theta} \in \mathbb{R}_+^3\}$ – o valor de $f(0, \alpha, \boldsymbol{\theta})$ não está definido, ainda que os limites sejam condizentes com os processos biológicos do TA:

$$\lim_{x \rightarrow \infty} f(x, \boldsymbol{\theta}) = 0 \quad \text{e} \quad \lim_{x \rightarrow 0^+} f(x, \boldsymbol{\theta}) = \beta_0$$

Quanto aos pontos críticos, verifica-se que primeira deverivada em relação a x é dada por:

$$\frac{\partial f}{\partial x} = \beta_1 e^{-\gamma x^\alpha} - \gamma \alpha (\beta_0 + \beta_1 x) e^{-\gamma x^\alpha} x^{\alpha-1},$$

cuja solução para x depende do valor arbitrário atribuído a α . A proposta dos autores Krewski et al. (1993) é $\alpha = 2$ e, então, tem-se que:

$$\left. \frac{\partial f}{\partial x} \right|_{\alpha=2} = 0 \Leftrightarrow \mathbf{x}^* = \left\{ x_1^* = \frac{-\gamma \beta_0 + \sqrt{(\gamma \beta_0)^2 + 2\gamma \beta_1^2}}{2\gamma \beta_1}, x_2^* = \frac{-\gamma \beta_0 - \sqrt{(\gamma \beta_0)^2 + 2\gamma \beta_1^2}}{2\gamma \beta_1} \right\}$$

É necessário observar que $x_2^* < 0$ e, dessa forma, não está no domínio de $f(x, 2, \boldsymbol{\theta})$. Além disso, como $\sqrt{(\gamma \beta_0)^2 + 2\gamma \beta_1^2} > \gamma \beta_0 \quad \forall \quad \gamma, \beta_0, \beta_1 \in \Theta$, então $f(x_1^*, 2, \boldsymbol{\theta})$ é um ponto de máximo mesmo para efeitos tóxicos muito mais evidentes que os efeitos mutagênicos, o que está ilustrado na Figura 56B.

Não há descontinuidades no preditor do modelo de Margolin (equação (5)) em todo o domínio $\{f(x, m, \boldsymbol{\theta}) : \mathbb{R}_+ \times \mathbb{Z}_+^* \times \mathbb{R}_+^3\}$. Para verificar os pontos críticos, obtém-se a

primeira derivada, em relação à dose x , que é dada por:

$$\frac{\partial f}{\partial x} = m\beta_1 e^{-(\beta_0 + \beta_1 x)} - m\gamma(1 - e^{-(\beta_0 + \beta_1 x)})e^{-\gamma x},$$

cujas soluções para x é

$$\frac{\partial f}{\partial x} = 0 \Leftrightarrow x^* = -\frac{\beta_0 + \log\left(\frac{\gamma}{\beta_1 + \gamma}\right)}{\beta_1}.$$

É fácil verificar que $f(x^*, m, \theta)$ é um ponto de máximo, desde que todos os parâmetros sejam positivos e, em caso contrário, não haverá ponto de máximo ou de mínimo. Esses comportamentos do preditor estão ilustrados na Figura 56C e é notável que quando $\gamma \gg \beta_1 \Rightarrow x^* \notin \mathbb{R}_+$.

Os limites de $f(x, m, \theta)$ respeitam os pressupostos biológicos, como segue:

$$\lim_{x \rightarrow \infty} f(x, \theta) = 0 \quad \text{e} \quad \lim_{x \rightarrow 0^+} f(x, \theta) = m(1 - e^{-\beta_0})$$

O preditor de Myers (equação (6)) é bem comportado em todos os aspectos: é contínuo em todo o domínio $\{f(x, \theta) : \mathbb{R}_+^4\}$, apresenta um único ponto crítico e os limites respeitam os pressupostos biológicos, como segue:

$$\frac{\partial f}{\partial x} = -e^{-\gamma x}(\gamma\beta_0 + \gamma\beta_1 x - \beta_1),$$

cujas soluções para x é

$$\frac{\partial f}{\partial x} = 0 \Leftrightarrow x^* = -\frac{\gamma\beta_0 - \beta_1}{\gamma\beta_1}.$$

$$\lim_{x \rightarrow \infty} f(x, \theta) = 0 \quad \text{e} \quad \lim_{x \rightarrow 0^+} f(x, \theta) = \beta_0$$

O ponto crítico no modelo de Myers é um ponto de máximo, mas quando $\gamma \gg \beta_1 \Rightarrow x^* \notin \mathbb{R}_+$, como pode ser observado na Figura 56D.

O preditor de Stead (equação (7)) não é contínuo em todo o domínio $\{f(x, \beta_0, \beta_1, \beta_2, \gamma) : x \geq 0, \beta_0 \geq 0, \beta_1 \geq 0, \beta_2 \geq 1, \gamma \geq 0\}$, pois $f(0, \cdot)$ não está definido. Os limites de f correspondem ao processo biológico, tal que:

$$\lim_{x \rightarrow \infty} f(x, \theta) = 0 \quad \text{e} \quad \lim_{x \rightarrow 0^+} f(x, \theta) = \beta_0$$

A determinação dos pontos críticos deste preditor exige a solução de uma equação não-linear:

$$\frac{\partial f}{\partial x} = \frac{\beta_1 \beta_2 x^{\beta_2} e^{-\gamma x}}{x} - \gamma(\beta_0 + \beta_1 x^{\beta_2})e^{-\gamma x},$$

cujas soluções para x envolve um polinômio de grau β_2 . Fixando $\beta_2 = 2$, pelo caráter ilustrativo, tem-se:

$$\left. \frac{\partial f}{\partial x} \right|_{\beta_2=2} = 0 \Leftrightarrow \mathbf{x}^* = \left\{ x_1^* = \frac{\beta_1 + \sqrt{\beta_1^2 - \gamma^2 \beta_1 \beta_0}}{\gamma \beta_1}, x_2^* = \frac{\beta_1 - \sqrt{\beta_1^2 - \gamma^2 \beta_1 \beta_0}}{\gamma \beta_1} \right\},$$

além da solução trivial. Para $\beta_2 = 2$, x_1^* e x_2^* são, respectivamente, pontos de máximo e mínimo, desde que $\beta_1 > \gamma^2 \beta_0$ e, em caso contrário, os pontos não são nem de máximo, nem de mínimo. No entanto, para $\beta_2 = 1$, há apenas um ponto de máximo, desde que $\beta_1 > \gamma^2 \beta_0$ e, em caso contrário, não há pontos de máximo ou mínimo. Como ilustrado na Figura 56E, para $\beta_2 = 1,5$ também existe apenas um ponto de máximo – os demais pontos críticos ou não são reais ou não estão no domínio.

Fazendo uma análise geral do comportamento dos preditores propostos, há uma questão interessante e que deve ser levada em conta na escolha do modelo estatístico: no preditor de Margolin, à medida em que o parâmetro associado à mutagenicidade aumenta (β_1) o máximo de $f(x^*, \cdot)$ é observado para valores cada vez menores de x^* , em contraste com todos os demais preditores. Assim, se esse modelo apresentar um bom ajuste aos dados, já é possível discutir acerca do comportamento mutagênico da substância em questão. O referido contraste entre os máximos dos preditores pode ser observado nas curvas em verde da Figura 56A–E. Se um ou mais desses modelos apresentarem bom ajuste, a dose de mutagênese máxima (DMM) pode ser obtida pela função `mmd.ames()`, descrita na seção 4.5.7 para cada modelo individualmente, ou combinar os resultados utilizando os pesos de Akaike, da equação (36).

Figura 56 - Flexibilidade dos preditores dos modelos tradicionalmente utilizados para análise do teste de *Salmonella*/microsoma. Os gráficos em A–E se referem respectivamente aos modelos de Breslow, Krewski, Margolin, Myers, e Stead. Cada padrão foi gerado por uma combinação diferente de parâmetros. Linha contínua em preto: referência para as demais curvas e se refere aos valores dos preditores obtidos com a esperança das distribuições uniformes para cada parâmetro da Tabela 7. Linhas tracejadas em cinza: efeito tóxico crescente (variação nos valores de γ). Linhas tracejadas em verde: efeito mutagênico crescente (variação nos valores de β_1). Linhas em azul: ausência de toxicidade ($\gamma = 0$). Linhas em vermelho: ausência de mutagenicidade ($\beta_1 = 0$). Linhas em alaranjado: efeito tóxico muito mais evidente do que o efeito mutagênico ($\gamma \gg \beta_1$), de forma que o máximo do preditor está fora do domínio. Linhas tracejadas em preto: variação no segundo parâmetro associado à mutagenicidade (β_2)