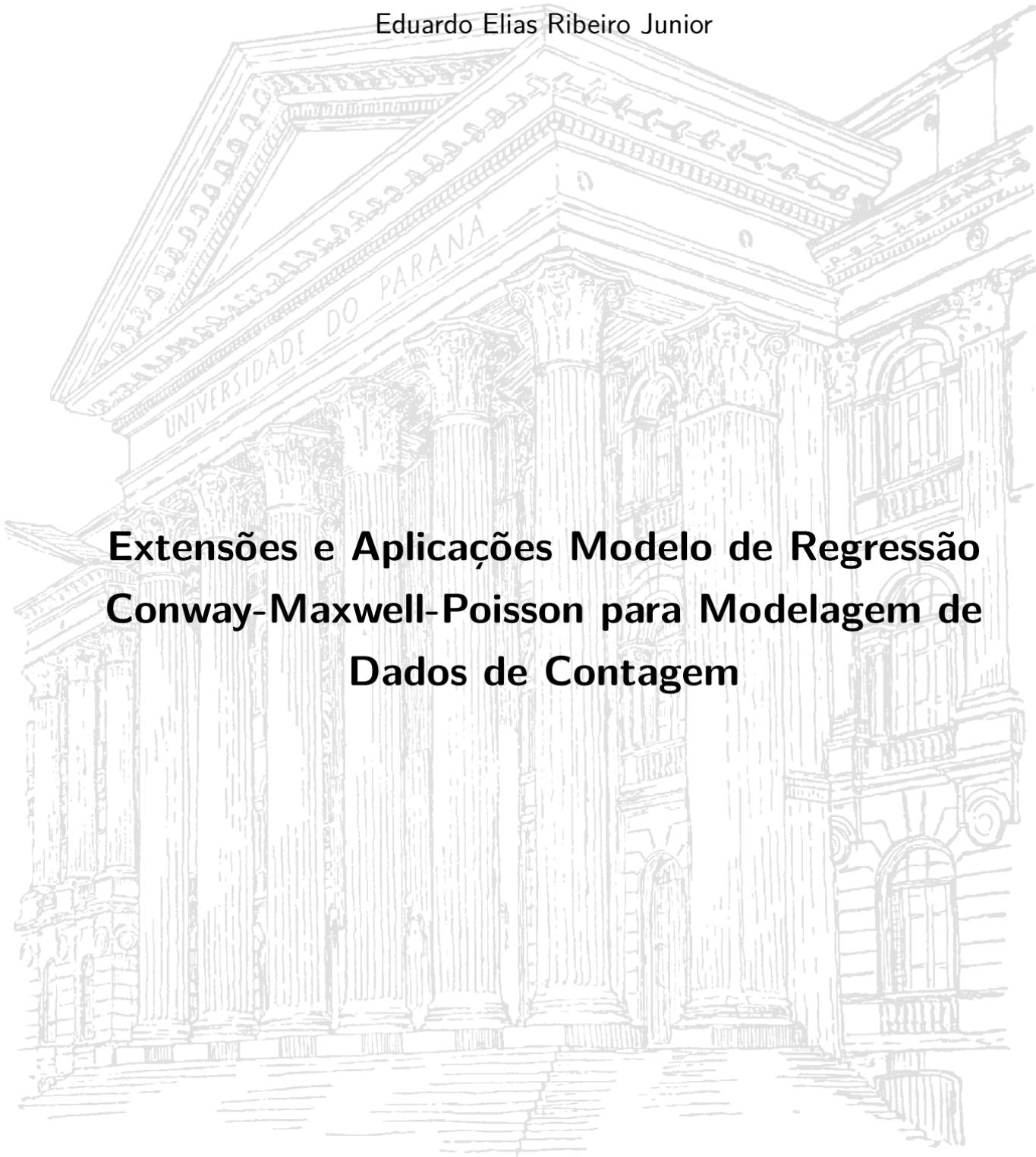


Universidade Federal do Paraná

Eduardo Elias Ribeiro Junior



**Extensões e Aplicações Modelo de Regressão
Conway-Maxwell-Poisson para Modelagem de
Dados de Contagem**

Curitiba

2016

Eduardo Elias Ribeiro Junior

**Extensões e Aplicações Modelo de Regressão
Conway-Maxwell-Poisson para Modelagem de Dados de
Contagem**

Trabalho de Conclusão de Curso apresentado à disciplina Laboratório B do Curso de Graduação em Estatística da Universidade Federal do Paraná, como exigência parcial para obtenção do grau de Bacharel em Estatística.

Universidade Federal do Paraná

Setor de Ciências Exatas

Departamento de Estatística

Orientador: Prof. Dr. Walmes Marques Zeviani

Curitiba

2016

Agradecimentos

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

*“Software is like sex: it’s better when
it’s free”*
— Linus Torvalds

*“The numbers are where the scientific
discussion should start, not end.”*
— Steven N. Goodman

Resumo

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Palavras-chave: COM-Poisson. Dados de contagem.

Lista de ilustrações

Figura 1 – Ilustração de diferentes tipos de processos pontuais	19
Figura 2 – Probabilidades pela distribuição Poisson para diferentes valores de	23
Figura 3 – Probabilidades pela distribuição Binomial Negativa para diferentes níveis de dispersão, fixando a média em 5	26
Figura 4 – Relação Média e Variância na distribuição Binomial Negativa	27
Figura 5 – Probabilidades pela distribuição COM-Poisson para diferentes parâmetros	28
Figura 6 – Exemplos de casos particulares da distribuição COM-Poisson	28
Figura 7 – Relação Média e Variância na distribuição COM-Poisson	29
Figura 8 – Convergência da constante de normalização da COM-Poisson para diferentes conjuntos de parâmetros	30
Figura 9 – Ilustração de dados de contagem com excesso de zeros	31
Figura 10 – (Esquerda) Número de capulhos produzidos para cada nível de desfolha e estágio fenológico	36
Figura 11 – Disposição das variáveis de contagem nº de estruturas reprodutivas, nº de capulhos produzidos e nº de nós da planta observadas sob diferentes dias de exposição à infestação de Mosca-branca	37
Figura 12 – Disposição das variáveis número de grãos e número de vagens nos diferentes níveis de adubação potássica e umidade do solo	38
Figura 13 – Médias e variâncias amostrais das contagens de grão e vagens, avaliadas no experimento com soja sob efeito umidade e adubação potássica	39
Figura 14 – (Esquerda) Dispersão entre o número total de ninfas de Mosca-branca nos folíolos da soja e o número de dias após a primeira avaliação para as quatro diferentes cultivares	40
Figura 15 – (Esquerda) Logarítmo neperiano do número de peixes capturados acrescido de 0,5 para as diferentes composições dos grupos	41
Figura 16 – Dispersão do número de nematóides em uma alíquota de 1ml provida pela solução de 1 g/ml de massa fresca diluída.	42
Figura 17 – Perfil de log-verossimilhança para o parâmetro extra da COM-Poisson, estimado no modelo com o quinto preditor	47
Figura 18 – Imagem da matriz de correlação entre os parâmetros do modelo COM-Poisson	48
Figura 19 – Curva dos valores preditos com intervalo de confiança de (95%) como função do nível de desfolha e do estágio fenológico da planta	49

Figura 20 – Perfis de log-verossimilhança para o parâmetro extra da COM-Poisson nos modelos para número de capulhos produzidos (esquerda), número de estruturas reprodutivas (central) e número de nós (direira) .	51
Figura 21 – Imagem da matriz de correlação entre os parâmetros do modelo COM-Poisson	52
Figura 23 – Curva dos valores preditos com intervalo de confiança de (95%) como função do nível de desfolha e do estágio fenológico da planta	52
Figura 22 – Imagem da matriz de correlação entre os parâmetros do modelo COM-Poisson	53
Figura 24 – Perfis de log-verossimilhança para o parâmetro extra da COM-Poisson nos modelos para número de vagens viáveis por parcela (esquerda) e número grãos de soja por parcela (direira)	55
Figura 25 – Imagem da matriz de correlação entre os parâmetros do modelo COM-Poisson ajustados ao número de vagens por parcela	56
Figura 26 – Imagem da matriz de correlação entre os parâmetros do modelo COM-Poisson ajustados ao número de grãos por parcela	57
Figura 27 – Valores preditos com intervalos de confiança (95%) como função do nível de adubação com potássio e do percentual de umidade do solo para cada variável de interesse mensurada (número de vagens e número de grãos por parcela)	58
Figura 29 – Imagem da matriz de correlação entre os parâmetros do modelo COM-Poisson	60
Figura 30 – Valores preditos com intervalos de confiança (95%) em função das cultivares de soja e da data de avaliação da planta	61
Figura 28 – Perfil de log-verossimilhança para o parâmetro extra da COM-Poisson	61
Figura 31 – Valores preditos do número de peixes capturados considerando o número de crianças e pessoas no grupo e a presença de um campista	65

Lista de tabelas

Tabela 1 – Distribuições de probabilidades para dados de contagem com indicação das características contempladas	22
Tabela 2 – Médias e variâncias amostras das contagens avaliadas no experimento de capulhos de algodão sob efeito de Mosca-Branca	37
Tabela 3 – Medidas de ajuste para avaliação e comparação entre preditores e modelos ajustados	46
Tabela 4 – Estimativas dos parâmetros e razões entre as estimativa e erro padrão para os três modelos em estudo	47
Tabela 5 – Medidas de ajuste para avaliação e comparação entre preditores e modelos ajustados	50
Tabela 6 – Medidas de ajuste para avaliação e comparação entre preditores e modelos ajustados	54
Tabela 7 – Medidas de ajuste para avaliação e comparação entre preditores e modelos ajustados	59
Tabela 8 – Medidas de ajuste para avaliação e comparação de preditores e modelos com componente de barreira ajustados	63
Tabela 9 – Estimativas dos parâmetros e razões entre as estimativa e erro padrão para os três modelos em estudo	64

Lista de símbolos

\log	Logarítmo neperiano (de base e).
ℓ	log-verossimilhança maximizada.
AIC	Critério de Informação de Akaike, do inglês <i>Akaike Information Criterion</i> .

Sumário

1	INTRODUÇÃO	17
2	MODELOS PARA DADOS DE CONTAGEM	21
2.1	Modelo Poisson	22
2.1.1	Estimação via Quase-Verossimilhança	24
2.2	Modelo Binomial Negativo	25
2.3	Modelo COM-Poisson	27
2.4	Modelos para excesso de zeros	31
2.5	Modelos de efeitos aleatórios	33
3	MATERIAL E MÉTODOS	35
3.1	Materias	35
3.1.1	Conjuntos de dados	35
3.1.1.1	Capulhos de algodão sob efeito de desfolha artificial	35
3.1.1.2	Produtividade de algodão sob efeito de infestação de Mosca-branca	36
3.1.1.3	Produtividade de soja sob efeito de umidade do solo e adubação potássica	38
3.1.1.4	Ocorrência de ninfas de Mosca-branca em lavoura de soja	39
3.1.1.5	Peixes Capturados por Pescadores em um Parque Estadual	39
3.1.1.6	Número de nematóides em raízes de feijoeiro	40
3.1.2	Recursos computacionais	41
3.2	Métodos	42
4	RESULTADOS E DISCUSSÃO	45
4.1	Análise de dados de capulhos de algodão sob efeito de desfolha	45
4.2	Análise de dados de capulhos de algodão sob efeito de Mosca-Branca	49
4.3	Análise de produção de soja sob efeito de umidade e adubação potássica	53
4.4	Análise de ninfas de mosca-branca em lavoura de soja	57
4.5	Análise de captura de peixes em um parque estadual	62
5	CONSIDERAÇÕES FINAIS	67
	REFERÊNCIAS	69

APÊNDICES	73
APÊNDICE A – LIPSUM	75
ANEXOS	77
ANEXO A – LIPSUM	79

1 Introdução

Em diversas áreas do conhecimento é comum o interesse em i) compreender o relacionamento entre variáveis de interesse e características de uma amostra e ii) realizar previsões por meio de modelos estatísticos ajustados por dados de uma amostra. A teoria de modelos de regressão sustentam muitas das pesquisas na área de Estatística aplicada.

Os modelos de regressão, na sua forma univariada e usual, consistem no estabelecimento de uma equação matemática que relaciona a média de uma variável aleatória de interesse (variável resposta) com as demais variáveis aleatórias observadas (covariáveis). Nesta metodologia considera-se uma distribuição de probabilidades para a variável resposta condicionada as covariáveis cuja a média está associada a um preditor que acomoda os efeitos das variáveis independentes.

Podemos destacar o modelo linear normal como o modelo predominante dentre as análises estatísticas aplicadas. Esse modelo estabelece que a variável resposta condicional as covariáveis têm distribuição Normal de média descrita por um preditor linear das covariáveis. Todavia, não são raras as situações em que a variável resposta se apresenta na forma de contagens, assumindo valores inteiros não negativos. Variáveis aleatórias de contagem, de forma geral, representam o número de ocorrências de um evento em um domínio específico que pode ser contínuo, como um intervalo de tempo ou espaço, ou discreto, como indivíduos ou grupos.

A análise de dados de contagem pelo modelo linear normal produz estimativas que contêm erros padrões inconsistentes e podem produzir previsões negativas para o número de eventos (KING, 1989). Uma alternativa adotada durante muitos anos, e ainda aplicada, é encontrar alguma forma de transformação da variável resposta a fim de atender aos pressupostos do modelo de regressão normal. Contudo essa abordagem dispõe de resultados insatisfatórios, pois i) dificulta a interpretação dos resultados, ii) não contempla a natureza da variável (ainda serão valores pontuais, só que em outra escala) iii) não contempla a relação média e variância, característica de dados de contagem e iv) no uso da transformação logarítmica é problemática quando há contagens nulas.

Diante do problema diferentes abordagens foram propostas, contudo destaca-se o trabalho apresentado por Nelder e Wedderburn (1972) que introduz a teoria dos modelos lineares generalizados (MLG's). Esta nova classe de modelos flexibilizou a distribuição condicional associada permitindo outras distribuições pertencentes à família exponencial de distribuições. Tal família contempla as distribuições Poisson, Binomial, Gama entre outras bem conhecidas na literatura, além da própria distribuição

Normal.

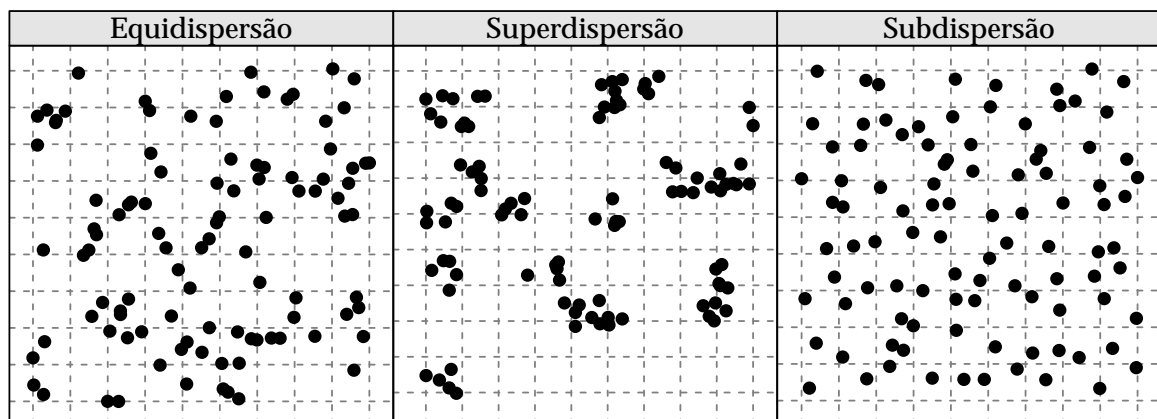
Com os MLG's a modelagem de dados passou a ser mais fiel a natureza da variável resposta, principalmente no que diz respeito ao seu suporte. Neste contexto, a análise de variáveis aleatórias de contagem, que têm suporte nos conjunto dos números naturais, foi enriquecida expressivamente.

Para análise estatística dessas variáveis, temos o modelo probabilístico de Poisson, já consolidado na literatura e amplamente utilizado. Este modelo possui apenas um parâmetro, denotado por λ , que representa a média e também a variância, o que implica em uma relação identidade ($\lambda = E[Y] = V[Y]$). Essa propriedade, chamada de equidispersão, é uma particularidade do modelo Poisson que pode não ser adequada a diversas situações. Quando aplicado sob negligência desta suposição, o modelo Poisson apresenta erros padrões inconsistentes para as estimativas dos parâmetros e por consequência, para toda função desses parâmetros (WINKELMANN, 1995; WINKELMANN; ZIMMERMANN, 1994).

O caso de superdispersão, quando a variância é maior que a média, é o mais comum e tem uma gama de métodos para análise mais extensa. A superdispersão pode ocorrer pela ausência de covariáveis importantes, excesso de zeros, diferentes amplitudes de domínio (*offset*) não consideradas, heterogeneidade de unidades amostrais, entre outros (Ribeiro Jr et al., 2012). Para tais casos uma abordagem é a adoção de modelos com efeitos aleatórios que capturam a variabilidade extra. Um caso particular dos modelos Poisson de efeitos aleatórios, muito adotado no campo aplicado da Estatística, ocorre quando consideramos distribuição Gama para os efeitos aleatórios, nesta situação temos expressão fechada para a função de probabilidade marginal, que assume a forma Binomial Negativa.

Outra manifestação de fuga da suposição de equidispersão é a subdispersão, situação menos comum na literatura. Os processos que reduzem a variabilidade das contagens, abaixo do estabelecido pela Poisson, não são tão conhecidos quanto os que produzem variabilidade extra. Pela mesma razão, são poucas as abordagens descritas na literatura que capazes de tratar a subdispersão, uma vez que efeitos aleatórios só capturam a variabilidade extra. Podemos citar os modelos de quasi-verossimilhança como a abordagem mais utilizada. Todavia não é possível recuperar a verdadeira distribuição da variável resposta nessa abordagem pois a modelagem é baseada apenas nos dois primeiros momentos da distribuição condicional (PAULA, 2013).

A figura 1 ilustra, sob um contexto espacial de duas dimensões, a ocorrência das características de equi, super e subdispersão respectivamente. Nesta figura cada ponto representa a ocorrência de uma variável aleatória e cada parcela, delimitada pelas linhas pontilhadas, representa o intervalo no espaço cujo contabiliza-se as ocorrências. No painel da esquerda temos a representação de dados de contagem equidispersos, neste



Fonte: Elaborado pelo autor.

Figura 1 – Ilustração de diferentes tipos de processos pontuais. Da direita para esquerda temos processos sob padrões aleatório, aglomerado e uniforme

cenário temos que as ocorrências da variável aleatória se dispõem aleatoriamente. No painel central o padrão já se altera, temos a representação do caso de superdispersão. Note que neste cenário formam-se aglomerados que deixam parcelas com contagens muito elevadas e parcelas com contagens baixas. Uma possível causa deste padrão se dá pelo processo de contágio (e.g. contagem de casos de uma doença contagiosa, contagem de frutos apodrecidos). Na terceiro e último painel temos o caso de subdispersão, em que as ocorrências se dispõem uniformemente no espaço. Note agora que as contagens de ocorrências nas parcelas variam bem pouco. Ao contrário do caso superdisperso uma causa provável seria o oposto de contágio, a repulsa, ou seja, uma ocorrência causa a repulsa de outras ocorrências em seu redor (e.g. contagem de árvores, contagem de animais).

Outra alternativa paramétrica que contempla os casos de equi, super e subdispersão é a adoção de uma distribuição mais flexível para a variável resposta condicional as covariáveis. A distribuição COM-Poisson surgiu anteriormente à formalização dos MLG's, proposta por [Conway e Maxwell \(1962\)](#) a COM-Poisson (nome em homenagem aos seus autores Richard W. Conway, William L. Maxwell, Conway-Maxwell-Poisson) generaliza a distribuição Poisson com a adição de mais um parâmetro, denotado por ν , que torna a razão de probabilidades sucessivas não linear contemplando os casos de sub e superdispersão ([SHMUELI et al., 2005](#)).

Uma característica bastante relevante é que a COM-Poisson possui como casos particulares as distribuições Poisson, Geométrica e Binomial. Portanto, empregando a COM-Poisson como distribuição condicional associada, obtemos um modelo de regressão sem a imposição de equidispersão. Tal flexibilidade, considerando o amplo uso do modelo Poisson, significa que a COM-Poisson pode ser aplicada nessas situações

e será especialmente importante naquelas onde há fuga da equidispersão.

Pela similaridade da função de distribuição COM-Poisson com a Poisson, vários aspectos podem ser estendidos. Por exemplo, há situações em que o delineamento do experimento sugere uma estrutura de covariância entre observações induzidas por um processo hierárquico de casualização ou amostragem. São casos assim os experimentos em parcelas subdivididas e experimentos com medidas repetidas ou longitudinais. Tais estruturas estabelecem modelos com efeitos não observáveis que agem no nível de observação ou unidade experimental e isso pode ser incorporado no modelo de regressão COM-Poisson com a inclusão de efeitos aleatórios. Da mesma forma, excesso de zeros pode ser introduzido a essa distribuição da mesma maneira que ocorre para o modelo Poisson, através de truncamento (modelos Hurdle) ou inflação (modelos de mistura) (SELLERS; RAIM, 2016). Estas extensões para o modelo COM-Poisson ainda não são bem consolidadas na literatura e são escassas suas aplicações. Uma constatação do fato é que não há implementações destas extensões nos principais softwares estatísticos.

Na literatura brasileira, aplicações do modelo COM-Poisson são escassas. Foram encontradas apenas aplicações na área de Análise de Sobrevida, mais especificamente em modelos com fração de cura (RIBEIRO, 2012; BORGES, 2012). Portanto, o presente trabalho visa colaborar com a literatura estatística brasileira i) apresentando e explorando o modelo de regressão COM-Poisson para dados de contagem, ii) estendendo as aplicações desse modelo COM-Poisson para situações específicas como inclusão de efeitos aleatórios e modelagem de excesso de zeros, iii) discutindo os aspectos inferenciais por meio de análise de dados reais e iv) disponibilizando os recursos computacionais, em formato de pacote R, para ajuste dos modelos apresentados. Nas aplicações optou-se também pela análise via modelos já disponíveis para as situações estudadas.

O trabalho é organizado em cinco capítulos. Esse primeiro capítulo visa enfatizar as características das variáveis aleatórias de contagem e suas lacunas que podem ser complementadas na análise estatística dessas variáveis. O capítulo 2 é dedicado a revisão bibliográfica dos modelos estatísticos empregados a análise de dados de contagem, nesse capítulo os modelos Poisson, Binomial Negativo, as abordagens para excesso de zeros, a estrutura dos modelos de efeitos aleatórios e o modelo COM-Poisson são apresentados. No capítulo ?? apresentamos os conjuntos de dados a serem analisados e os métodos para ajuste e comparação dos modelos. O capítulo 4 traz os principais resultados da aplicação e comparação dos modelos estatísticos com ênfase nas discussões sob aspectos inferenciais empíricos. Finalmente no capítulo 5 são apresentadas as considerações finais obtidas desse trabalho e listados algumas possíveis linhas de pesquisa para estudos futuros.

2 Modelos para dados de contagem

Métodos para inferência em dados de contagem estão bem aquém da quantidade disponível para dados contínuos. Destacamos o modelo log-linear Poisson como o modelo mais utilizado quando se trata de dados de contagem. Porém não raramente os dados de contagens apresentam variância superior ou inferior à sua média. Esses são os casos de super ou subdispersão já enunciados no capítulo 1, que quando ocorrem inviabilizam o uso da distribuição Poisson.

Nos casos de fuga da equidispersão algumas abordagens não paramétricas são empregadas. Nesse contexto, podemos citar os métodos de estimação via quase-verossimilhança, estimação robusta dos erros padrões (estimador “sanduíche”) e estimação dos erros padrões via reamostragem (“*bootstrap*”) (HILBE, 2014). Desses métodos detalhamos brevemente somente o método de estimação via função de quase-verossimilhança na seção 2.1.1.

No contexto paramétrico, pesquisas recentes trazem modelos bastante flexíveis à fuga de equidispersão no campo da Estatística aplicada, veja (SELLERS; SHMUELI, 2010; ZEVIANI et al., 2014; LORD; GEEDIPALLY; GUIKEMA, 2010). Na tabela 1 listamos as distribuições de probabilidades consideradas por Winkelmann (2008) e Kokonendji (2014) e as características de dados de contagem que são contempladas. Notamos que a Poisson na verdade é um caso particular, pois é a única das distribuições listada que contempla somente a característica de equidispersão, ainda observa-se que temos um conjunto maior de distribuições para os casos de superdispersão com relação os casos de subdispersão. Embora este grande número de distribuições exista para lidar com os casos de fuga de equidispersão destacamos que são poucos os pacotes estatísticos que empregam essas distribuições a modelos de regressão para dados de contagem.

Dos modelos paramétricos o Binomial Negativo aparece em destaque com implementações já consolidadas nos principais *softwares* estatísticos e frequentes aplicações nos casos de superdispersão. Na seção 2.2 detalhes da construção desses modelos são apresentados. Dos demais modelos derivados das distribuições listadas na tabela ?? este trabalho abordará somente o modelo COM-Poisson, que é apresentado com detalhes na seção 2.3.

Um outro fenômeno que é frequente em dados de contagem é a ocorrência excessiva de zeros. Esse fenômeno sugere a modelagem de dois processos geradores de dados, o gerador de zeros extra e o gerador das contagens. Existem ao menos duas abordagens pertinentes para estes casos que são os modelos de mistura e os modelos condicionais. Na abordagem por modelos de mistura a variável resposta é modelada

Tabela 1 – Distribuições de probabilidades para dados de contagem com indicação das características contempladas

Distribuição	Contempla a característica de		
	Equidispersão	Superdispersão	Subdispersão
Poisson	✓		
Binomial Negativa	✓	✓	
<i>Inverse Gaussian Poisson</i>	✓	✓	
<i>Compound Poisson</i>	✓	✓	
Poisson Generalizada	✓	✓	✓
<i>Gamma-Count</i>	✓	✓	✓
COM-Poisson	✓	✓	✓
Katz	✓	✓	✓
<i>Poisson Polynomial</i>	✓	✓	✓
<i>Double-Poisson</i>	✓	✓	✓
<i>Lagrangian Poisson</i>	✓	✓	✓

Fonte: Elaborado pelo autor.

como uma mistura de duas distribuições, no trabalho de [Lambert \(1992\)](#), uma mistura da distribuição Bernoulli com uma distribuição de Poisson ou Binomial Negativa. Considerando os modelos condicionais, também chamados de modelos de barreira ([RIDOUT; DEMETRIO; HINDE, 1998](#)), temos que a modelagem da variável resposta é realizada em duas etapas. A primeira refere-se ao processo gerador de contagens nulas e a segunda ao gerador de contagens não nulas. Nesta trabalho a modelagem de excesso de zeros se dará somente via modelos de barreira. A seção 2.4 é destinada a um breve detalhamento desta abordagem.

Nesta capítulo também abordamos a situação da inclusão de efeitos aleatórios no seção 2.5. Em análise de dados de contagem a inclusão desses efeitos permitem acomodar variabilidade extra e incorporar a estrutura amostral do problema como em experimentos com medidas repetidas ou longitudinais e experimentos em parcelas subdivididas.

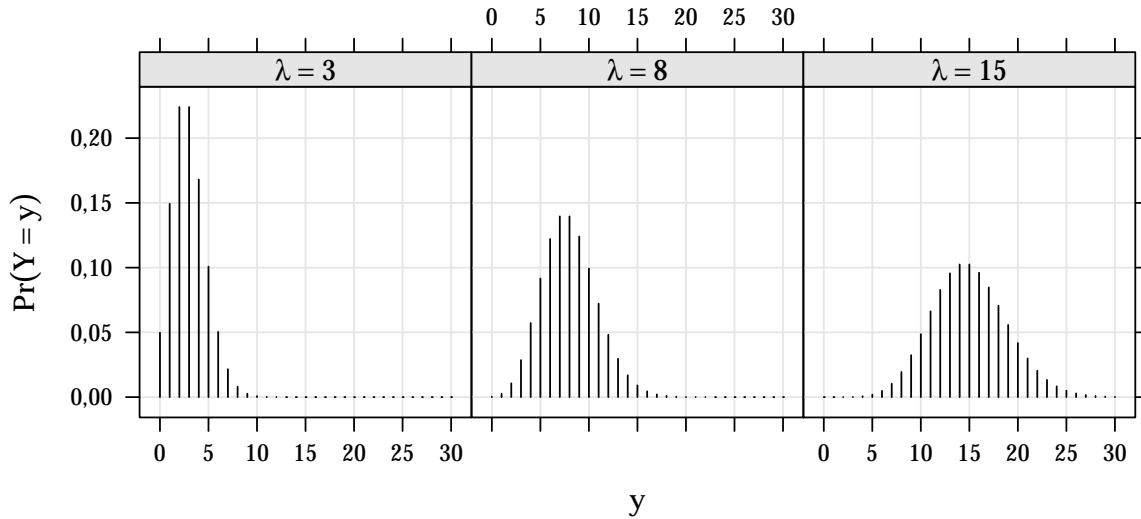
2.1 Modelo Poisson

A Poisson é uma das principais distribuição de probabilidades discretas. Com suporte nos inteiros não negativos, dizemos que uma variável aleatória segue um modelo Poisson se sua função massa de probabilidade for

$$Pr(Y = y | \lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \quad y = 0, 1, 2, \dots \quad (2.1)$$

em que $\lambda > 0$ representa a taxa de ocorrência do evento de interesse. Uma particulari-

dade já destacada desta distribuição é que $E(X) = V(X) = \lambda$. Isso torna a distribuição Poisson bastante reestrutiva. Na figura 2 são apresentadas as ditribuições Poisson para diferentes parâmetros, note que devido a propriedade $E(X) = V(X)$ contagens maiores também são mais dispersas.



Fonte: Elaborado pelo autor.

Figura 2 – Probabilidades pela distribuição Poisson para diferentes valores de λ

Uma propriedade importante da distribuição Poisson é sua relação com a distribuição Exponencial. Essa relação estabelece que se os tempos entre a ocorrência de eventos se distribuem conforme modelo Exponencial de parâmetro λ a contagem de eventos em um intervalo de tempo t tem distribuição Poisson com média λt . A distribuição *Gamma-Count*, citada na tabela 1, estende esta propriedade do processo adotando a distribuição Gama para os tempos entre eventos tornando a distribuição da contagem decorrente mais flexível (WINKELMANN, 1995; ZEVIANI et al., 2014).

Outra propriedade que decorre da construção do modelo Poisson é sobre a razão entre probabilidades sucessivas, $\frac{P(Y=y-1)}{P(Y=y)} = \frac{y}{\lambda}$. Essa razão é linear em y e tem sua taxa de crescimento ou decrescimento como $\frac{1}{\lambda}$. Os modelos Katz e COM-Poisson se baseiam na generalização da razão de probabilidades a fim de flexibilizar a distribuição decorrente.

A utilização do modelo Poisson na análise de dados se dá por meio do modelo de regressão Poisson. Seja Y_i variáveis aleatórias condicionalmente independentes, dados as covariáveis X_i , $i = 1, 2, \dots, n$. O modelo de regressão log-linear Poisson, sob a teoria

dos MLG's é definido como

$$\begin{aligned} Y_i | X_i &\sim \text{Poisson}(\mu_i) \\ \log(\mu_i) &= X_i\beta \end{aligned} \quad (2.2)$$

em que $\mu_i > 0$ é a média da variável aleatória $Y_i | X_i$ que é calculada a partir do vetor $\beta \in \mathbb{R}^p$.

O processo de estimação do vetor β é baseado na maximização da verossimilhança que nas distribuições que pertencem à família exponencial, os MLG's, é realizado via algoritmo de mínimos quadrados ponderados iterativamente, ou, do inglês *Iterative Weighted Least Squares* - IWLS (NELDER; WEDDERBURN, 1972).

2.1.1 Estimação via Quase-Verossimilhança

Em 1974 Wedderburn propôs uma forma de estimação a partir de uma função bi-paramétrica, denominada quase-verossimilhança. Suponha que temos y_i observações independentes com esperanças μ_i e variâncias $V(\mu_i)$. A função de quase-verossimilhança é expressa como

$$Q(\mu_i | y_i) = \int_y^{\mu_i} \frac{y_i - t}{\sigma^2 V(\mu_i)} dt \quad (2.3)$$

Na expressão 2.3 a função de quase-verossimilhança é definida a partir da especificação de μ_i , $V(\mu_i)$ e σ^2 . O processo de estimação via maximização dessa função compartilha as mesmas estimativas para μ_i , porém a dispersão de y_i , $V(y_i) = \theta V(\mu_i)$ é corrigida pelo parâmetro adicional σ^2 .

Assim os problemas com a fuga da suposição de equidispersão podem ser superados quando a estimação por máxima quase-verossimilhança é adotado. Porém um resultado dessa abordagem é que

$$-E \left(\frac{\partial^2 Q(\mu | y)}{\partial \mu^2} \right) \leq -E \left(\frac{\partial^2 \ell(\mu | y)}{\partial \mu^2} \right) \quad (2.4)$$

ou seja a informação a respeito de μ quando se conhece apenas σ^2 e $V(\mu)$, a relação entre média e variância, é menor do que a informação quando se conhece a distribuição da variável resposta, dada pela log-verossimilhança $\ell(\mu | y)$. Além disso ressalta-se que, de forma geral, não se recupera a distribuição de Y somente com as especificações de σ^2 e $V(\mu)$.

Em modelos de regressão, definimos $g(\mu_i) = X\beta$ e $V(\mu_i)$ que definem a função de quase-verossimilhança. Nessa abordagem são estimados os parâmetros β e σ^2 . A

estimativa do vetor β pode ser obtidas pelo algoritmo *IWLS*. Usando as funções quase-escore e matriz de quase-informação chega-se ao mesmo algoritmo de estimação dado no caso Poisson, que não depende de σ^2 . O parâmetro σ^2 é estimado separadamente, pós estimação dos β 's. Um estimador usual é o baseado na estatística χ^2 de Pearson.

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \quad (2.5)$$

2.2 Modelo Binomial Negativo

Uma das principais alternativas paramétricas para dados de contagem superdispersos é a adoção da distribuição Binomial Negativa. A função massa de probabilidade da distribuição Binomial Negativa pode ser deduzida de um processo hierárquico de efeitos aleatórios onde se assume que

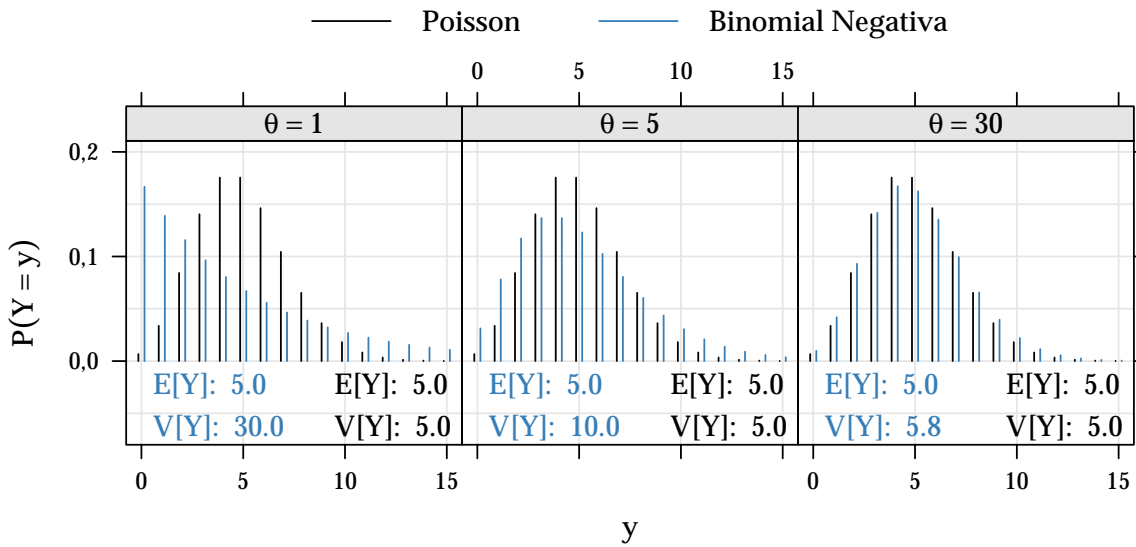
$$\begin{aligned} Y | b &\sim \text{Poisson}(b) \\ b &\sim \text{Gama}(\mu, \theta) \end{aligned} \quad (2.6)$$

A função massa de probabilidade decorrente da estrutura descrita em 2.7 é deduzida integrando os efeitos aleatórios. Considere $f(y | b)$ como a função massa de probabilidade da distribuição Poisson (vide expressão em 2.1) e $g(b | \mu, \theta)$ a função densidade da distribuição Gama ¹

$$\begin{aligned} \Pr(Y = y | \mu, \theta) &= \int_0^\infty f(y | b) g(b | \mu, \theta) db \\ &= \frac{\theta^\theta}{y! \mu^\theta \Gamma(\theta)} \int_0^\infty e^{-b(1+\theta/\mu)} b^{y+\theta-1} db \\ &= \frac{\Gamma(\theta + y)}{\Gamma(y+1) \Gamma(\theta)} \left(\frac{\mu}{\mu + \theta} \right)^y \left(\frac{\theta}{\mu + \theta} \right)^\theta \quad y = 0, 1, 2, \dots \end{aligned} \quad (2.7)$$

com $\mu > 0$ e $\theta > 0$. Ressaltamos que esse é um caso particular de um modelo de efeito aleatório cuja a integral tem solução analítica e por consequência o modelo marginal tem forma fechada. Outro caso que se baseia no mesmo princípio é o modelo *Inverse Gaussian Poisson*, que como o nome sugere adota a distribuição Inversa Gaussiana para os efeitos aleatórios. Na figura 3 são apresentadas as distribuições Binomial Negativa para diferentes parâmetros θ em comparação com a distribuição Poisson equivalente em locação. Note que quanto menor o parâmetro θ , maior a dispersão da distribuição. Isso introduz uma propriedade importante desse modelo, para $\theta \rightarrow \infty$ a distribuição reduz-se a Poisson.

¹ O desenvolvimento detalhado da integral pode ser visto em Paula (2013, pág. 303-305). Obs.: A função densidade do modelo Gama está parametrizada para que μ represente a média da distribuição.



Fonte: Elaborado pelo autor.

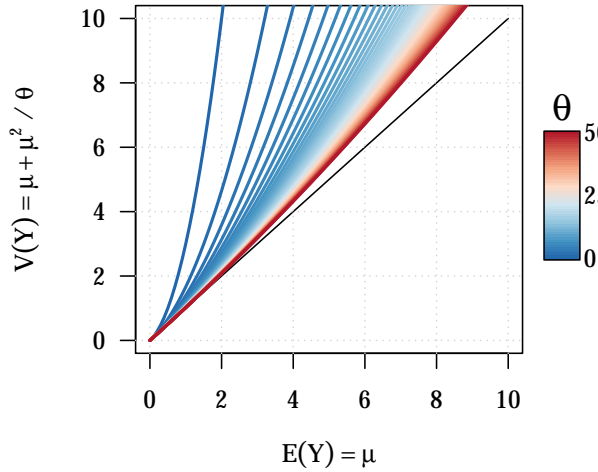
Figura 3 – Probabilidades pela distribuição Binomial Negativa para diferentes níveis de dispersão, fixando a média em 5.

Os momentos média e variância da distribuição Binomial Negativa são expressos como $E(Y) = \mu$ e $V(Y) = \mu + \mu^2/\sigma^2$. Note que pelas expressões fica evidente a característica da Binomial Negativa de acomodar somente superdispersão, pois $E(Y)$ é menor que $V(Y)$ para qualquer σ^2 . Percebemos também quanto maior o parâmetro σ^2 mais $E(Y)$ se aproxima de $V(Y)$, e no limite, quando $\sigma^2 \rightarrow \infty$, $E(Y) = V(Y)$ fazendo com que a distribuição Binomial Negativa se reduza a Poisson.

A relação funcional entre média e variância é ilustrada na figura 4 onde apresentamos as médias e variâncias para μ entre 0 e 10 e θ entre 0 e 50. O comportamento dessa relação proporciona um maior flexibilidade à distribuição em acomodar superdispersão, uma característica importante exibida nesta figura é que para a Binomial Negativa se aproximar a Poisson em contagens altas o θ deve ser extremamente grande.

O emprego do modelo Binomial Negativo em problemas de regressão ocorre de maneira similar aos MLG's, com exceção de que a distribuição só pertence a família exponencial de distribuições se o parâmetro θ for conhecido e assim o processo sofre algumas alterações. Primeiramente, assim como na Poisson, definimos $g(\mu_i) = X\beta$, comumente utiliza-se a função $g(\mu_i) = \log(\mu_i)$. Desenvolvendo a log-verossimilhança e suas funções derivadas, função score e matriz de informação de Fisher chegamos que a matriz de informação é bloco diagonal caracterizando a ortogonalidade dos parâmetros β de locação e θ de dispersão. Deste fato decorre que a estimação dos parâmetros pode ser realizada em paralelo, ou seja, estima-se o vetor β pelo método de IWLS e posteriormente o parâmetro θ pelo método de Newton-Raphson, faz-se os dois procedimentos simultaneamente até a convergência dos parâmetros.

2.3 Modelo COM-Poisson



Fonte: Elaborado pelo autor.

Figura 4 – Relação Média e Variância na distribuição Binomial Negativa

A distribuição de probabilidades COM-Poisson foi proposta em 1962, em um contexto de filas por [Conway e Maxwell](#) e generaliza a Poisson em termos da razão de probabilidades sucessivas, como veremos adiante. Seja Y uma variável aleatória COM-Poisson, então sua função massa de probabilidade é

$$Pr(Y = y \mid \lambda, \nu) = \frac{\lambda^y}{(y!)^\nu Z(\lambda, \nu)} \quad y = 0, 1, 2, \dots \quad (2.8)$$

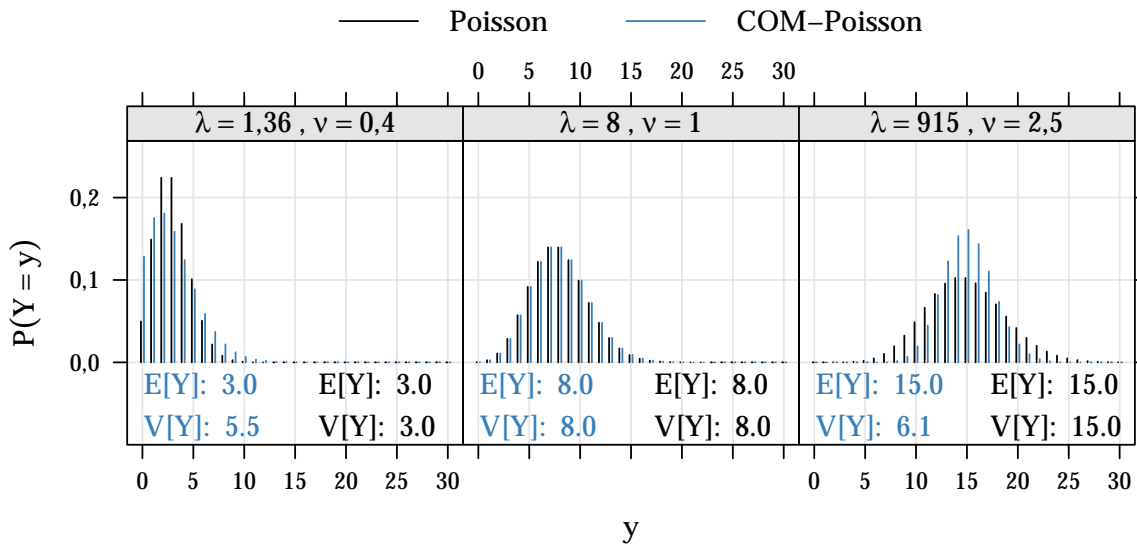
em que $\lambda > 0$, $\nu \geq 0$ e $Z(\lambda, \nu)$ é uma constante de normalização, calculada para que de fato seja uma função massa de probabilidade. $\sum_{i=1}^{\infty} Pr(Y = y) = 1$. $Z(\lambda, \nu)$ é definida como se segue

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu} \quad (2.9)$$

O fato que torna a distribuição COM-Poisson mais flexível é a razão entre probabilidades sucessivas

$$\frac{Pr(Y = y - 1)}{Pr(Y = y)} = \frac{y^\nu}{\lambda} \quad (2.10)$$

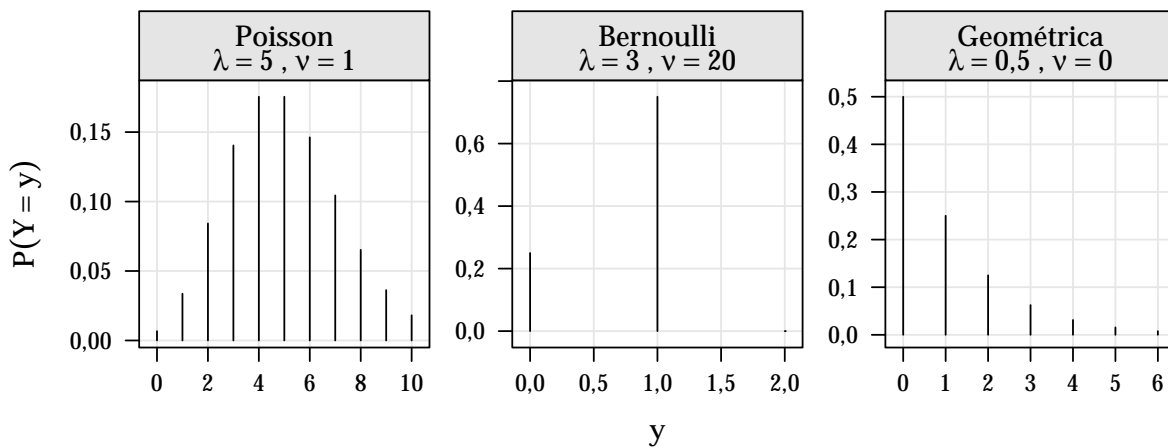
que se caracteriza não necessariamente linear em y , diferentemente da Poisson, o que permite caudas mais pesadas ou mais magras à distribuição ([SELLERS; SHMUELI, 2010](#)). Na figura 5 apresentamos as dsitribuições COM-Poisson para diferentes valores de λ e ν em contraste com as equivalentes, em locação, distribuições Poisson. Nessa figura podemos apreciar a flexibilidade desse modelo, pois i) contempla o caso de subdispersão mesmo em contagens baixas ($E(Y) = 3$, painel a esquerda), a distribuição permite caudas pesadas e conseqüentemente uma dispersão extra Poisson, ii) contempla subdispersão mesmo em contagens altas, o que na Poisson teríamos variabilidade na mesma magnitude, na COM-Poisson podemos ter caudas mais magras concentrando as probabilidades em torno da média (painel a direita) e iii) tem como caso particular a Poisson quando o parâmetro $\nu = 1$ (painel central).



Fonte: Elaborado pelo autor.

Figura 5 – Probabilidades pela distribuição COM-Poisson para diferentes parâmetros

Uma das vantagens do modelo COM-Poisson é que possui, além da Poisson quando $\nu = 1$ outras distribuições bem conhecidas como casos particulares. Esses casos particulares ocorrem essencialmente devido a forma assumida pela série infinita $Z(\lambda, \nu)$. Quando $\lambda = 1$, $Z(\lambda, \nu = 1) = e^\lambda$ e substituindo na expressão 2.8 temos a distribuição Poisson resultante. Quando $\nu \rightarrow \infty$, $Z(\lambda, \nu) \rightarrow 1 + \lambda$ e a distribuição COM-Poisson se aproxima de uma distribuição Bernoulli com $P(Y = 1) = \frac{\lambda}{1+\lambda}$. E quando $\nu = 0$ e $\lambda < 1$ $Z(\lambda, \nu)$ é uma soma geométrica que resulta em $(1 - \lambda)^{-1}$ e a expressão 2.8 se resume a uma distribuição Geométrica com $P(Y = 0) = (1 - \lambda)$ (SHMUELI et al., 2005). Os três respectivos casos particulares citados são ilustrados na figura 6, onde determinamos os parâmetros conforme restrições para redução da distribuição.



Fonte: Elaborado pelo autor.

Figura 6 – Exemplos de casos particulares da distribuição COM-Poisson

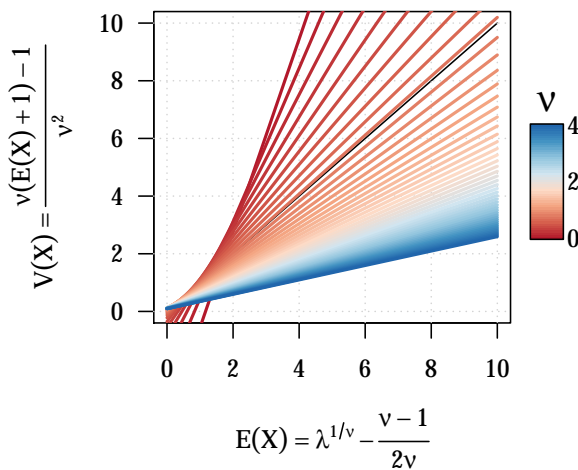
Um inconveniente desse modelo é que os momentos média e variância não tem forma fechada. Sendo assim podem ser calculados a partir da definição

$$E(Y) = \sum_{y=0}^{\infty} y \cdot p(y) \quad V(Y) = \sum_{y=0}^{\infty} y^2 \cdot p(y) - E^2(Y)$$

[Shmueli et al. \(2005\)](#), a partir de uma aproximação para $Z(\lambda, \nu)$, apresenta uma forma aproximada para os momentos da distribuição

$$E(Y) \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu} \quad V(Y) \approx \frac{\lambda^{1/\nu}}{\nu} \quad (2.11)$$

os autores ressaltam que essa aproximação é satisfatória para $\nu \leq 1$ ou $\lambda > 10^\nu$. Na figura 7 representamos de forma gráfica a relação média e variância aproximada pelas expressões em 2.11. Note que temos quase uma relação linear entre média e variância, [Sellers e Shmueli \(2010\)](#) descrevem que essa pode ser aproximada por $\frac{1}{\nu}E(Y)$. Dessas aproximações, bem como das visualizações em 5, 6 e 7 temos que o parâmetros ν , ou $\frac{1}{\nu}$, controla a precisão da distribuição, sendo ela equidispersa $\mu = 1$, superdispersa quando $\nu < 1$ e subdispersa quando $\nu > 1$.



Fonte: Elaborado pelo autor.

Figura 7 – Relação Média e Variância na distribuição COM-Poisson

Embora o modelo COM-Poisson não tenha expressão fechada para a média da distribuição pode-se utilizá-lo como modelo associado a distribuição condicional da variável resposta de contagem. Isso é feito incorporando um preditor linear em λ , que embora não representa a média está associado com a locação da distribuição, ou seja, modela-se a média indiretamente nessa abordagem. O modelo de regressão é definido com as variáveis aleatórias condicionalmente independentes Y_1, Y_2, \dots, Y_n , dado o vetor de covariáveis $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ seguindo um modelo COM-Poisson de parâmetros $\lambda_i = e^{X_i\beta}$, $i = 1, 2, \dots, n$ e ν comum a todas as observações. Sob a notação de MLG's, temos em 2.2 o modelo devida-

mente formulado

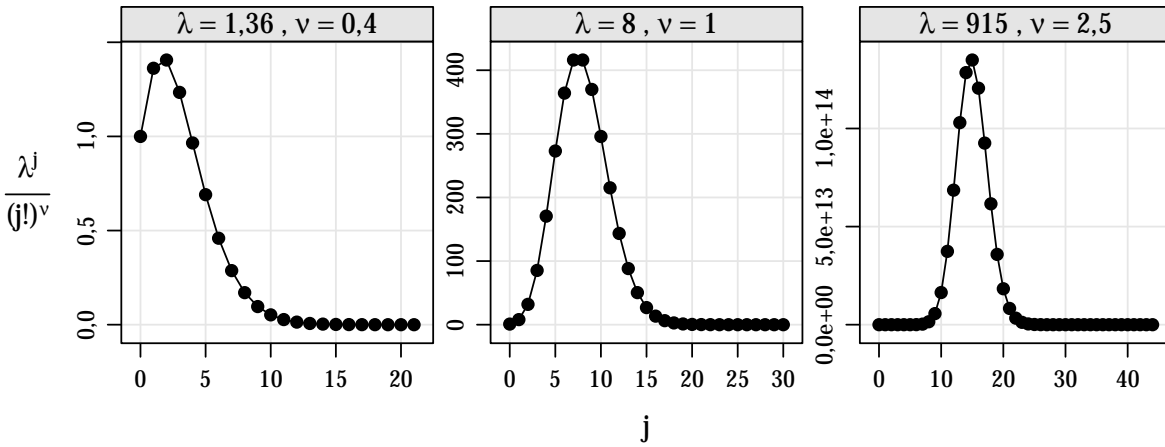
$$\begin{aligned} Y_i | X_i &\sim \text{COM-Poisson}(\lambda_i, \nu) \\ \eta(E(Y_i | X_i)) &= \log(\lambda_i) = X_i \beta \end{aligned} \quad (2.12)$$

O algoritmo para estimação do conjunto de parâmetros $\Theta = (\nu, \beta)$ do modelo é baseado na maximização da log-verossimilhança, que decorrente da especificação em 2.12 é

$$\ell(\nu, \beta | \underline{y}) = \sum_i^n y_i \log(\lambda_i) - \nu \sum_i^n \log(y_i!) - \sum_i^n \log(Z(\lambda_i, \nu)) \quad (2.13)$$

e então as estimativas de máxima verossimilhança são

$$\hat{\Theta} = (\hat{\nu}, \hat{\beta}) = \arg \max_{(\nu, \beta)} \ell(\nu, \beta | \underline{y})$$



Fonte: Elaborado pelo autor.

Figura 8 – Convergência da constante de normalização da COM-Poisson para diferentes conjuntos de parâmetros

Note que nessa maximização a constante de normalização $Z(\lambda, \nu)$, conforme definida em 2.9 é calcula para cada indivíduo o que potencialmente torna o processo de estimação lento. Uma ilustração do número de incrementos considerados para cálculo da constante $Z(\lambda, \nu)$ é apresentado na figura 8, neste ilustração foram utilizados os mesmos parâmetros definidos em 5 e note que o número de incrementos considerados para convergência². de $Z(\lambda, \nu)$ foram 22, 31, 45 nos primeiro, segundo e terceiro painéis respectivamente.

² Adotou-se como critério de convergência a iteração j tal que $\lambda^j / (j!)^\nu < 0,00001$

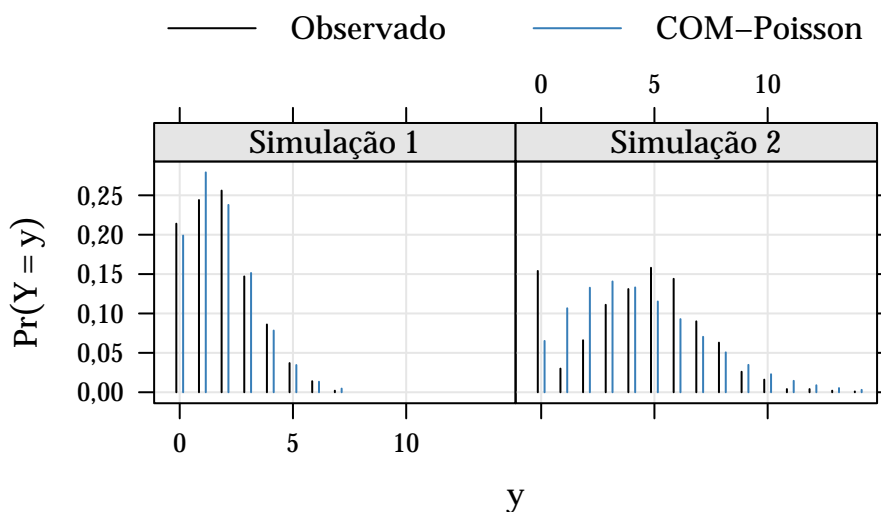
Detalhes computacionais do algoritmo de maximização e manipulações algébricas para eficiência na avaliação da log-verossimilhança no modelo COM-Poisson são discutidos na seção 3.2.

2.4 Modelos para excesso de zeros

Problemas com excesso de zeros são comuns em dados de contagem. Caracteriza-se como excesso de zeros casos em que a quantidade observada de contagens nulas supera substancialmente aquela esperada pelo modelo de contagem adotado, no caso do modelo Poisson $e^{-\lambda}$.

As contagens nulas que geram o excesso de zeros podem ser decorridas de duas formas distintas. A primeira denominamos de zeros estruturais, quando a ocorrência de zero se dá pela ausência de determinada característica na população e a segunda, que denominamos zeros amostrais ocorre segundo um processo gerador de dados de contagem (e.g processo Poisson). Assim, de forma geral temos dois processos geradores de dados atuantes na geração de uma variável aleatória de contagem com excessivos zeros.

Em geral, quando dados de contagem apresentam excessos de valores zero também apresentarão subdispersão. Todavia, essa dispersão pode ser exclusivamente devido ao excesso de zeros e assim os modelos alternativos já apresentados não terão um bom desempenho. Uma ilustração deste fato é ilustrada pela figura 9, em que simulamos um conjunto de dados com excesso de ajustamos um modelo COM-Poisson. Note que em ambos os casos o modelo se ajustou adequadamente, indicando os excessos de zeros devem ser abordados de forma diferente.



Fonte: Elaborado pelo autor.

Figura 9 – Ilustração de dados de contagem com excesso de zeros

Hilbe (2014, capítulo 7) discute sobre a interpretação e modelagem de dados de contagem com excesso de zeros. Para essa situação temos ao menos duas abordagens i) os modelos de mistura (LAMBERT, 1992), também chamados de inflacionados, em inglês *Zero Inflated Models* e ii) os modelos condicionais (RIDOUT; DEMETRIO; HINDE, 1998), também chamados de modelos de barreira, em inglês *Hurdle Models*. Neste trabalho somente a abordagem via modelos condicionais será abordada. A função massa de probabilidade do modelo Hurdle é

$$Pr(Y = y \mid \pi, \Theta_c) = \begin{cases} \pi & \text{se } y = 0, \\ (1 - \pi) \frac{Pr(Z = z \mid \Theta_c)}{1 - Pr(Z = 0 \mid \Theta_c)} & \text{se } y = 1, 2, \dots \end{cases} \quad (2.14)$$

em que $0 < \pi < 1$, representa a probabilidade de ocorrência de zeros e $Pr(Z = z \mid \Theta_c)$ a função massa de probabilidade de uma variável aleatória de contagem Z , como a Poisson ou a Binomial Negativa.

Da especificação em 2.14, os momentos média e variância são obtidos facilmente usando as definições $E(Y) = \sum_{y=1}^{\infty} y \cdot Pr(Y = y)$ e $V(Y) = \sum_{y=1}^{\infty} y^2 \cdot Pr(Y = y) - E^2(Y)$

$$E(Y) = \frac{E(Z)(1 - \pi)}{1 - Pr(Z = 0)} \quad V(Y) = \frac{1 - \pi}{1 - Pr(Z = 0)} \left[E(Z) \frac{(1 - \pi)}{1 - Pr(Z = 0)} \right]$$

Para a inclusão de covariáveis, caracterizando um problema de regressão, dado que o modelo tem dois processos atuantes devemos modelar ambos como se segue

$$\log \left(\frac{\pi_i}{1 - \pi_i} \right) = G_i \gamma \quad e \quad \begin{aligned} Z_i &\sim D(\mu_i, \phi) \\ g(\mu_i) &= X_i \beta \end{aligned} \quad (2.15)$$

com $i = 1, 2, \dots, n$, G_i e X_i as covariáveis da i -ésima observação consideradas para explicação da contagens nulas e não nulas respectivamente, $D(\mu_i, \phi)$ uma distribuição de probabilidades para considerada para as contagens não nulas que pode conter ou não um parâmetro ϕ adicional, se Poisson $D(\mu_i, \phi)$ se resume a $Poisson(\mu_i)$ e $g(\mu_i)$ uma função de ligação, nos casos Poisson e Binomial Negativa considera-se $\log(\mu_i)$. O que está implícito na formulação 2.15 é que para a componente que explica a geração de zeros está sendo considerada a distribuição Bernoulli de parâmetro π_i , contudo pode-se utilizar distribuições censuradas a direita no ponto $y = 1$ para estimação desta probabilidade, como explicam Zeileis, Kleiber e Jackman (2007).

2.5 Modelos de efeitos aleatórios

Nas seções anteriores exploramos modelos que flexibilizam algumas suposições do modelo Poisson. Basicamente pertimindo casos não equidispersos e modelando conjuntamente um processo gerador de zeros extra. Contudo uma suposição dos modelos de regressão para dados de contagem vistos até aqui é que as variáveis aleatória Y_1, Y_2, \dots, Y_n são condicionalmente independentes, dado o vetor de covariáveis. Porém não são raras as situações em que essa suposição não se mostra adequada. [Ribeiro \(2012\)](#) cita alguns exemplos:

- as observações podem ser correlacionadas no espaço,
- as observações podem ser correlacionadas no tempo,
- interações complexas podem ser necessárias para modelar o efeito conjunto de algumas covariáveis,
- heterogeneidade entre indivíduos ou unidades podem não ser suficientemente descrita por covariáveis.

Nessas situações pode-se estender a classe de modelos de regressão com a adição de efeitos aleatórios que incorporam variáveis não observáveis (latentes) ao modelo, permitindo assim acomodar uma variabilidade, que pode ser ou não estruturada, não prescrita pelo modelo. De forma geral a especificação dos modelos de efeitos aleatórios segue uma especificação hierárquica

$$\begin{aligned} Y_{ij} \mid b_i, X_{ij} &\sim D(\mu_{ij}, \phi) \\ g(\mu_{ij}) &= X_{ij}\beta + Z_i b_i \\ b &\sim K(\Theta_b) \end{aligned} \tag{2.16}$$

para $i = 1, 2, \dots, m$ (grupos com efeitos aleatórios comuns) e $j = 1, 2, \dots, n$ (observações) com $D(\mu_{ij}, \phi)$, uma distribuição considerada para as variáveis resposta condicionalmente independentes, $g(\mu_{ij})$ uma função de ligação conforme definida na teoria dos MLG's, X_{ij} e Z_i as vetores conhecidos representando os efeitos das covariáveis de interesse, b_i uma quantidade aleatória provida de uma distribuição $K(\Theta_b)$. Note que nesses modelos uma quantidade aleatória é somada ao preditor linear, diferentemente dos modelos de efeitos fixos e a partir desta quantidade é possível induzir um comportamento correlato entre as observações.

Como temos duas quantidades aleatórias no modelo, $Y | X$ e b , a verossimilhança para um modelo de efeito aleatório é dada integrando-se os efeitos aleatórios

$$\mathcal{L}(\beta, \phi, \Theta_b | \underline{y}) = \prod_{i=1}^m \int_{\mathbb{R}^q} \left(\prod_{j=1}^{n_i} f_D(y_{ij}, \mu, b_i) \right) \cdot f_K(b | \Theta_b) db_i \quad (2.17)$$

Perceba que na avaliação da verossimilhança é necessário o cálculo de m integrais de dimensão q . Para muitos casos essa integral não tem forma analítica sendo necessários métodos numéricos de aproximação, que são discutidos na seção 3.2. E as estimativas de máxima verossimilhança são

$$\hat{\Theta} = (\hat{\beta}, \hat{\Theta}_b) = \arg \max_{(\beta, \Theta_b)} \log(\mathcal{L}(\beta, \phi, \Theta_b | \underline{y}))$$

note que no processo de estimação dos modelos de efeitos aleatórios, métodos numéricos são intensivamente utilizados, pois a cada iteração do algoritmo de maximização da log-verossimilhança m integrais de dimensão q são aproximadas, ou seja, métodos de aproximação de integrais são utilizados concomitantemente ao método de maximização.

Em modelos de contagem de efeitos mistos é comum adotar como distribuição para os efeitos aleatórios uma Normal q -variada com média 0 e matriz de variância e covariâncias Σ , ou seja, na especificação 2.16 $K(\Theta_b) = NMV_q(0, \Sigma)$. Para estes casos os principais métodos de aproximação da integral tem desempenhos melhores (BATES et al., 2015).

Como mencionado anteriormente modelos de efeitos aleatórios são candidatos a modelagem de dados superdispersos. Quando não há uma estrutura de delineamento experimental ou observacional pode-se incluir efeitos aleatórios a nível de observação (e então $m = n$, ou seja, os vetores Y e b tem mesma dimensão). Casos particulares de modelos de efeitos aleatórios, onde o efeito aleatório é adiciona a nível de observação são o modelo Binomial Negativo e o *Inverse Gaussian Model*, em ambos os casos a integral, definida em 2.17 tem solução analítica e consequentemente a marginal em Y forma fechada.

3 Material e Métodos

Essa seção é destinada a apresentação dos conjuntos de dados analisados no trabalho e descrição dos recursos computacionais e métodos utilizados na análise. Na seção 3.1.1 os conjuntos de dados são apresentados, ao todo são seis conjuntos de dados com diferentes características. Os recursos computacionais utilizados são descritos na seção 3.1.2. Na última seção 3.2 desse capítulo são apresentados os métodos para ajuste, avaliação e comparação dos modelos propostos.

3.1 Materias

3.1.1 Conjuntos de dados

A seguir são apresentados os seis conjuntos de dados utilizados para avaliar o desempenho dos modelos COM-Poisson. Os dados em estudo são, quase em sua totalidade, resultantes de experimentos agronômicos com delineamentos balanceados, o que é uma característica desejável, haja visto que o principal interesse é a avaliação do desempenho do modelo COM-Poisson quando empregado a análise desses dados.

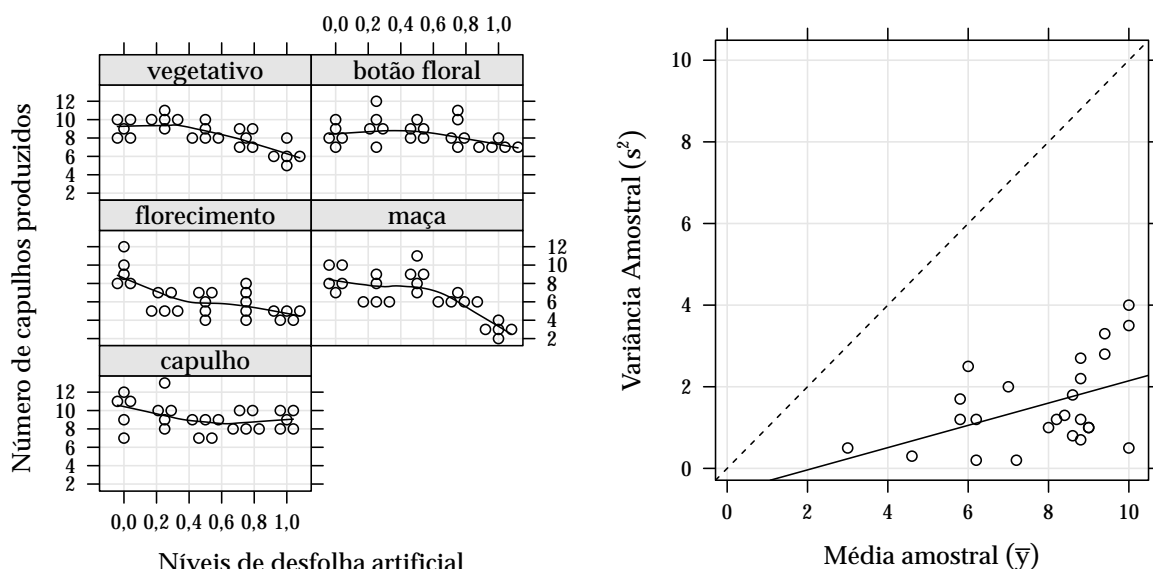
A apresentação dos conjuntos segue a ordem de 1) descrição do experimento ou estudo em destaque, 2) definição das variáveis e suas unidades de medidas e 3) descrição de suas características, potencialmente contempladas por modelos alternativos ao Poisson.

3.1.1.1 Capulhos de algodão sob efeito de desfolha artificial

Experimento conduzido sob delineamento inteiramente casualizado com cinco repetições em casa de vegetação com plantas de algodão *Gossypium hirsutum* submetidas à diferentes níveis de desfolha artificial de remoção foliar (0, 25, 50, 75, 100%), em combinação com o estágio fenológico no qual a desfolha foi aplicada (vegetativo, botão floral, florecimento, maça, capulho). A unidade experimental foi um vaso com duas plantas onde avaliou-se o número de capulhos produzidos ao final da ciclo cultura (ZEVIANI et al., 2014 apud SILVA et al., 2012). O experimento contou com 125 observações das quais temos as informações das variáveis número de capulhos de algodão produzidos, ncap, nível de desfolha de remoção foliar, des e estágio fenológico das planta na unidade experimental, est.

Esse conjunto de dados já fora publicado sob a motivação da característica de subdispersão, na ocasião o modelo proposto na análise foi o *Gamma-Count* (ZEVIANI et al., 2014). Na figura 10, apresentamos os dados do experimento. A esquerda temos a

disposição das cinco observações em cada tratamento (combinação de nível de desfolha e estágio fenológico do algodão) e à direita um gráfico descritivo cruzando médias e variâncias amostrais calculadas em cada tratamento, onde a linha pontilhada representa a característica de equidispersão, média igual a variância. Em todos os tratamentos obteve-se a média menor que a variância apontando evidência de subdispersão.



Fonte: Traduzido de Zeviani et al. (Figura 2)

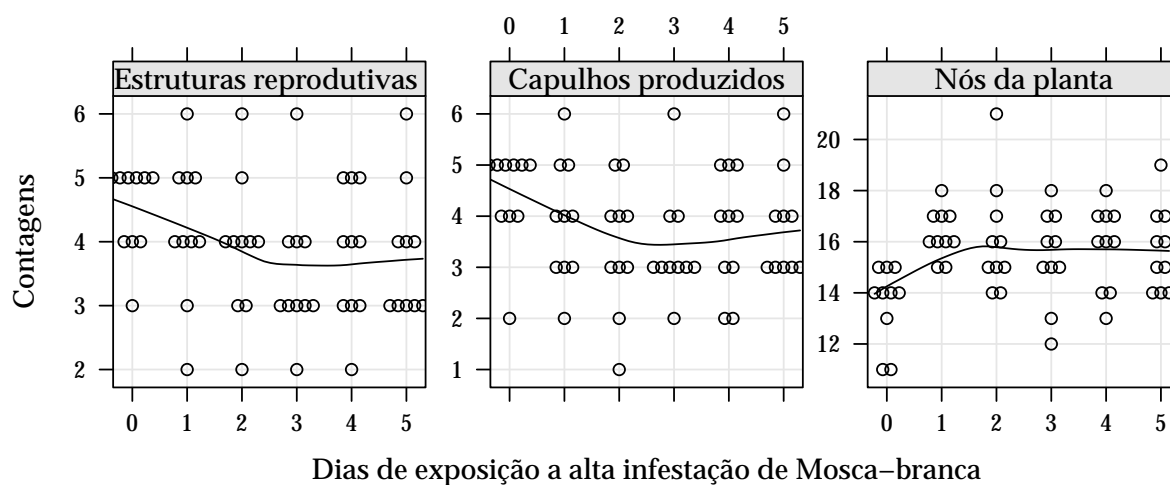
Figura 10 – (Esquerda) Número de capulhos produzidos para cada nível de desfolha e estágio fenológico. (Direita) Médias e variâncias das cinco repetições em cada combinação de nível de desfolha e estágio fenológico.

3.1.1.2 Produtividade de algodão sob efeito de infestação de Mosca-branca

Experimento conduzido na Universidade Federal da Grande Dourados (UFGD) em 2007, cujo objetivo foi avaliar os impactos da exposição de plantas à alta infestação de Mosca-branca *Bemisia tabaci* em componentes de produção do algodão¹. No experimento, plantas de algodão foram expostas à alta infestação da praga por diferentes períodos, 0, 1, 2, 3, 4, 5 dias onde avaliou-se o número de capulhos produzidos, ncapu, o número de estruturas reprodutivas, nrep e o número de nós nnos, como variáveis de interesse que representam a produtividade do cultivo de algodão. A condução do estudo deu-se via delineamento inteiramente casualizado com cinco vasos contendo duas plantas, para cada período de exposição.

Na figura 11 apresentamos a disposição de cada uma das variáveis aleatórias de contagem número de estruturas reprodutivas, número de capulhos produzidos e número de nós da planta para os diferentes períodos em que as plantas estiveram sob alta

¹ Experimento ainda não publicado.



Fonte: Elaborado pelo autor.

Figura 11 – Disposição das variáveis de contagem n° de estruturas reprodutivas, n° de capulhos produzidos e n° de nós da planta observadas sob diferentes dias de exposição à infestação de Mosca-branca.

infestação de Mosca-branca. Para todas as contagens parece haver um comportamento subdisperso. A indicação de subdispersão também se observa na tabela 2, onde temos as médias e variâncias amostrais calculadas com as dez observações nos seis períodos de exposição à infestação de Mosca-branca. Em todos os casos observa-se as variâncias amostrais substancialmente menores que respectivas médias, ainda a manifestação de subdispersão é mais expressiva na variável número de nós da planta. Portanto, nesse experimento modelos alternativos ao Poisson devem ser empregados, pois a suposição de equidispersão é violada.

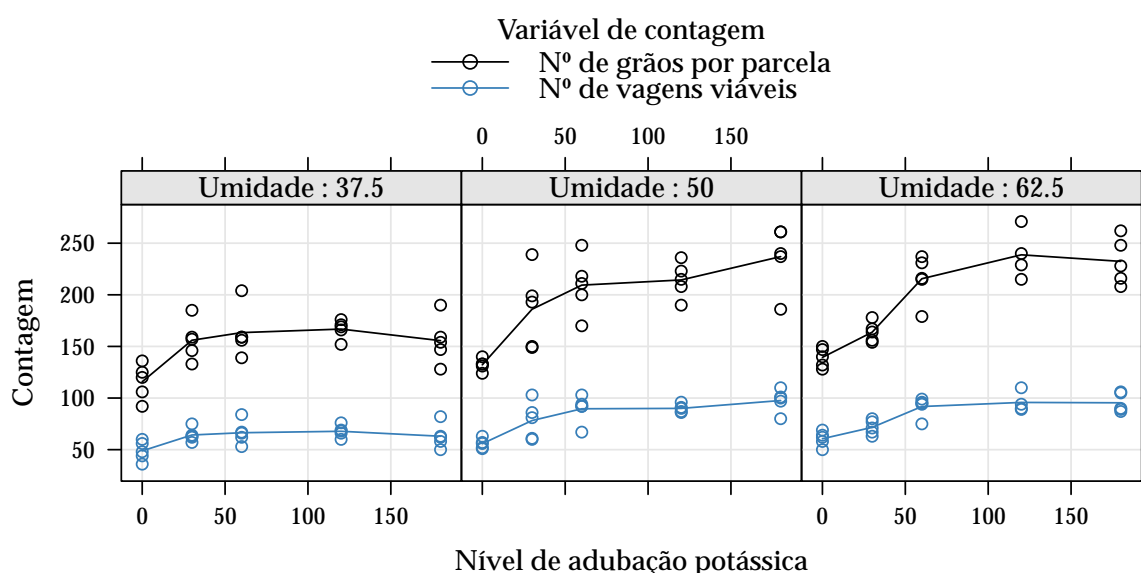
Tabela 2 – Médias e variâncias amostras das contagens avaliadas no experimento de capulhos de algodão sob efeito de Mosca-Branca

Dias de Exposição	N. Estruturas		N. Capulhos		N. Nós	
	média	variância	média	variância	média	variância
0	4,50	0,50	4,40	0,93	13,60	2,27
1	4,20	1,29	3,90	1,43	16,30	0,90
2	3,90	1,21	3,40	1,60	16,10	4,54
3	3,50	1,17	3,40	1,16	15,40	3,38
4	3,80	1,07	3,70	1,34	15,80	2,62
5	3,80	1,07	3,80	1,07	15,70	2,68

Fonte: Elaborado pelo autor.

3.1.1.3 Produtividade de soja sob efeito de umidade do solo e adubação potássica

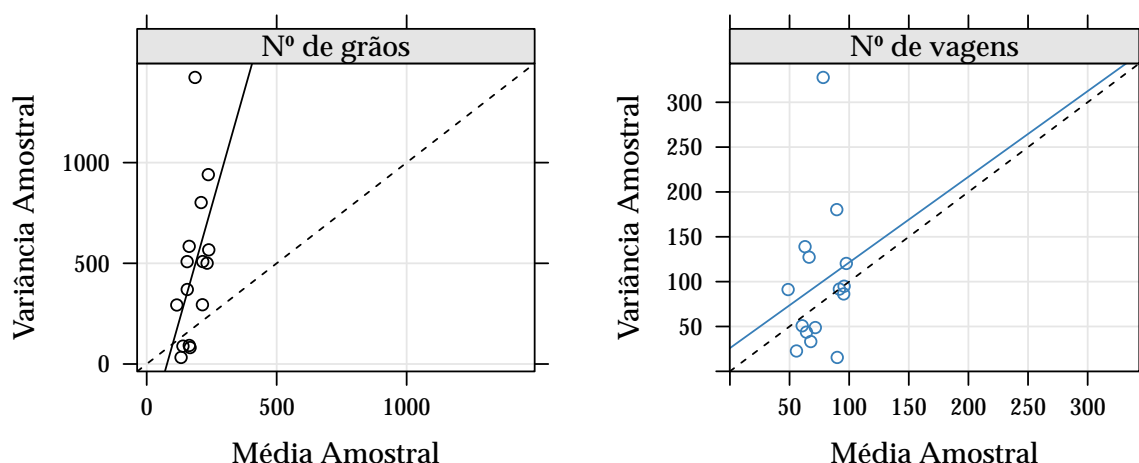
Experimento fatorial 5×3 que estudou diferentes níveis de adubação potássica aplicada ao solo, 0, 30, 60, 120, 180 mg dm^{-3} e diferentes níveis de umidade do solo, 37.5, 50, 62.5 % do volume total dos poros, que representam pouca água, água em quantidade ideal e água em abundância, nos componentes de produção da soja. O experimento foi instalado em casa de vegetação no delineamento de blocos casualizados completos e a unidade experimental foi um vaso com duas plantas de soja. No experimento foram medidas várias variáveis respostas (que representam a produtividade), sendo que o número de vagens viáveis por vaso e o número de grãos por vaso foram as variáveis de contagem (SERAFIM et al., 2012).



Fonte: Elaborado pelo autor.

Figura 12 – Disposição das variáveis número de grãos e número de vagens nos diferentes níveis de adubação potássica e umidade do solo.

Na figura 12 apresentamos a dispersão das contagens nas combinações das covariáveis umidade do solo e adubação potássica. As duas variáveis de contagem avaliadas no experimento apresentam níveis de dispersão distintos, essa característica fica explícita na figura 13, em que apresentamos as dispersões entre médias e variâncias amostrais para cada uma das variáveis. Para o número de grãos por parcela, com contagens mais elevadas, as variâncias amostrais são, quase em sua totalidade, superiores as médias caracterizando uma evidência de superdispersão. Já para o número de vagens por parcela temos médias e variâncias, em média, próximas o que é indício que a suposição de equidispersão é razoável.



Fonte: Elaborado pelo autor.

Figura 13 – Médias e variâncias amostrais das contagens de grão e vagens, avaliadas no experimento com soja sob efeito umidade e adubação potássica.

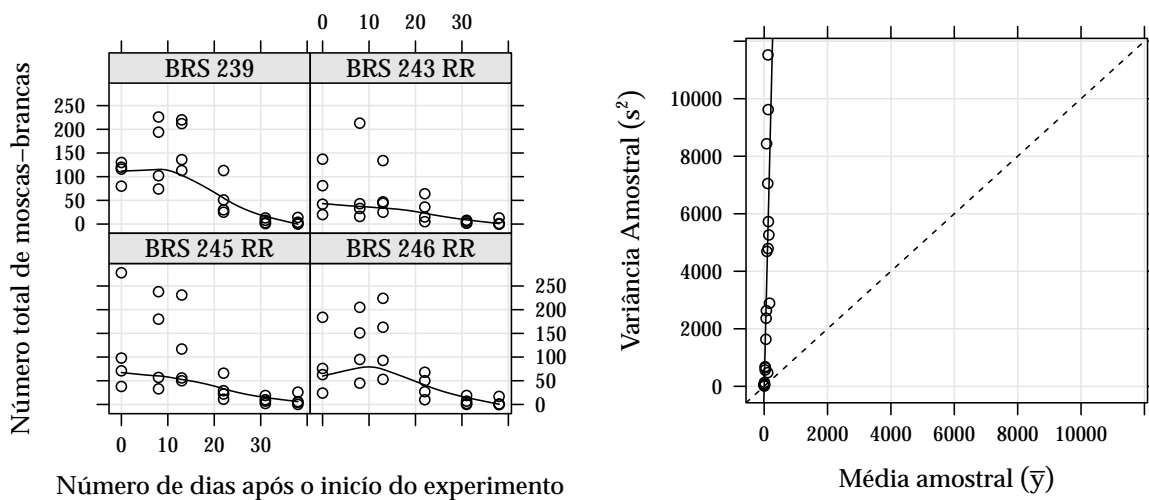
3.1.1.4 Ocorrência de ninfas de Mosca-branca em lavoura de soja

Nesse experimento também envolvendo a cultura de soja e a praga Mosca-branca, foram avaliadas plantas de quatro diferentes cultivares de soja, BRS 245 RR, BRS 243 RR, BRS 246 RR, BRS 239, contabilizando o número de ninfas de mosca-branca nos folíolos dos terços superior, médio e inferior das plantas em seis datas, 11/12/09, 19/12/09, 24/12/09, 02/01/10, 11/01/10, 18/01/10 dentre os 38 dias de estudo. O experimento foi conduzido em casa de vegetação sob o delineamento de blocos casualizados para controle de variação local (SUEKANE, 2011).

Na figura 14 a esquerda, apresentamos as contagens da praga para cada cultivar em cada uma das datas de avaliação, representadas pelos dias decorridos após a primeira avaliação, em 11/12/09. Note que as contagens são muito altas e dispersas, principalmente nas quatro primeiras avaliações. A direita temos uma descrição no nível de dispersão da variável de contagem. Perceba que esse é um conjunto de dados extremamente superdisperso, os pontos, que representam em médias e variâncias em cada combinação de dias de após a primeira avaliação e cultivares de soja, estão todos acima da reta identidade (equidispersão) com variâncias em torno de 1.000 vezes maiores que as respectivas médias.

3.1.1.5 Peixes Capturados por Pescadores em um Parque Estadual

Diferentemente dos demais, esse é um estudo observacional feito por biólogos com interesse em modelar o número de peixes capturados por grupos de pescadores visitantes em um Parque Estadual **citar o repositório da UCLA**. Nesse estudo tem-se



Fonte: Elaborado pelo autor

Figura 14 – (Esquerda) Dispersão entre o número total de ninfas de Mosca-branca nos folíolos da soja e o número de dias após a primeira avaliação para as quatro diferentes cultivares. (Direita) Relação entre as médias e as variâncias amostrais do número de ninfas nesse experimento.

como informações a respeito dos grupos de visitantes, o número de pessoas e de crianças no grupo e se há ou não a presença de campista. Um fato interessante deste dado é que nem todos os grupos de visitantes praticaram pescaria, portanto, nesses grupos o número de peixes capturado será zero.

Na figura 15 pode-se notar o evidente excesso de contagens zero. No gráfico a esquerda apresentamos a disposição das contagens, transformadas por $\log(y_i|x_i + 0,5)$. É característica marcante no gráfico a grande quantidade de pontos dispostos no primeiro valor do eixo y , $-0,693 = \log(0.5)$. Embora seja um gráfico marginal, não considerando as covariáveis de cada contagem, a direita temos um histograma da variável resposta onde percebe-se novamente a grande quantidade de valores nulos, ao todo 56,8% dos dados são contagens nulas. Portanto nesse problema, claramente modelos alternativos que acomodem excesso de zeros se fazem necessários.

3.1.1.6 Número de nematóides em raízes de feijoeiro

Esse último conjunto de dados explorado no trabalho, é resultado de um experimento em casa de vegetação que estudou a reprodução de nematóides em cultivares de feijoeiro. No experimento, o solo de vasos com duas plantas de feijão foi inicialmente contaminado com nematóides. Ao final do experimento, as raízes das duas plantas por vaso foram lavadas, trituradas, peneiradas e diluídas e as contagens dos nematóides realizadas em alíquotas dessa solução. Perceba que temos, incidindo sobre a contagem a quantidade de raízes produzidas por parcela e ainda o volume usado para diluir essa

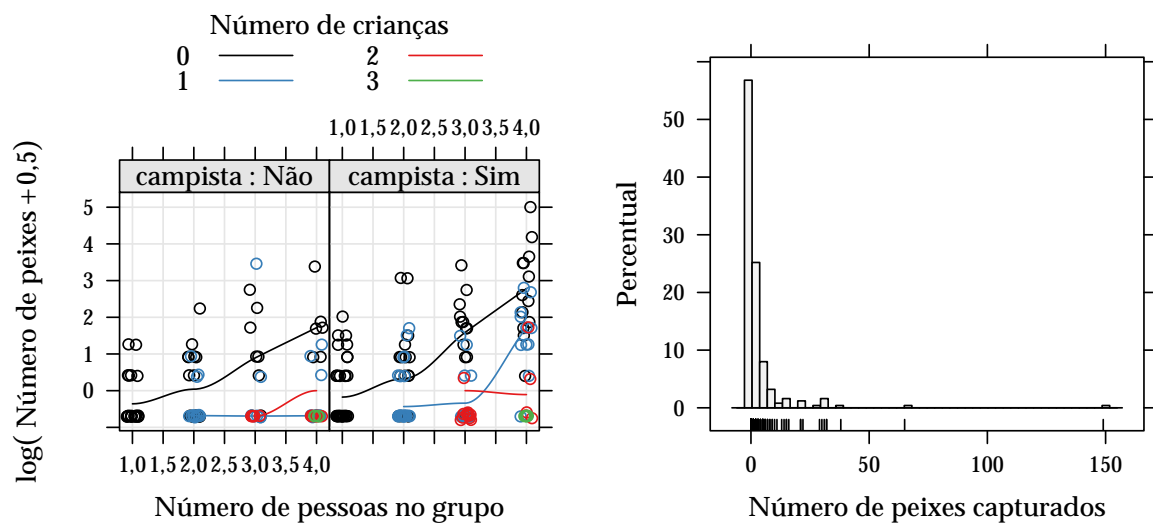


Figura 15 – (Esquerda) Logarítmo neperiano do número de peixes capturados acrescido de 0,5 para as diferentes composições dos grupos. (Direita) Histograma do número de peixes capturados por grupo.

quantidade, pois as contagens são realizadas em alíquotas da solução composta pela razão entre massa fresca de raízes (em gramas) e volume utilizado para diluição (em milímetros)².

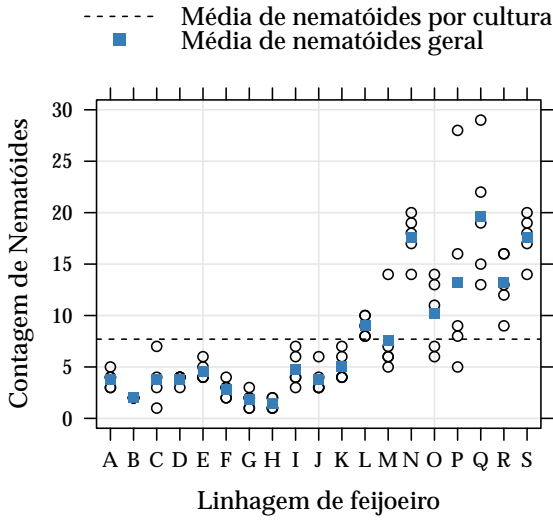
Na figura 16 apresentamos a dispersão das contagens de nematóides em uma alíquota de 1 ml da solução composta de uma grama de massa fresca de raiz por um milímetro de água. Note que as contagens para cada uma das culturas se distribuem em torno do perfil médio (linha pontilhada). Um detalhe interessante desse conjunto de dados é que as linhagens de feijoeiro avaliadas não foram escolhas subjetivas, pode-se interpretar as linhagens escolhidas como um sorteio aleatório dentre uma população de linhagens de feijoeiro. Assim esse experimento pode ser rotulado como um experimento de medidas repetidas, caso comum nas aplicações de modelos de efeitos aleatórios.

3.1.2 Recursos computacionais

O *software* R, versão 3.3.0, é utilizado tanto para a preparação e apresentação dos dados quanto para ajuste dos modelos e apresentação de resultados. Pacotes auxiliares utilizados no trabalho são: MASS (versão 7.3.45) para ajuste e inferências dos modelos Binomial Negativo, bbmle (versão 1.0.18) para estimação via máxima verossimilhança das funções implementadas para o modelo COM-Poisson, psc1 (versão 1.4.9) para ajuste dos modelos Poisson e Binomial Negativo com componente de barreira para modelagem de excesso de zeros e lme4 (versão 1.1.12) para ajuste dos modelos Poisson

² Cedido para fins acadêmicos por Andressa Cristina Zamboni Machado, pesquisadora do Instituto Agrônomo do Paraná (IAPAR), e pelo técnico agrícola do IAPAR Santino Aleandro da Silva

com efeitos aleatórios normais. Para apresentação gráfica dos resultados os pacotes `lattice` (versão 0.20.33), `latticeExtra` (versão 0.6.28) e `corrplot` (versão 0.73) são exaustivamente utilizados. Finalmente, para elaboração do relatório, mesclando códigos em R e escrita na linguagem de marcação \LaTeX , o pacote `knitr` (versão 1.12.3) é requerido.



Fonte: Elaborado pelo autor.

Figura 16 – Dispersão do número de nematóides em uma alíquota de 1ml provida pela solução de 1 g/ml de massa fresca diluída.

Destaca-se nesse trabalho que todas as funções implementadas para ajuste e inferência dos modelos de regressão COM-Poisson estão disponíveis, em formato de um pacote R `tccPackage`, no endereço <https://github.com/jreduardo/tccPackage>. No capítulo 4 destacamos as chamadas de algumas funções e no apêndice A mostramos o emprego do pacote na análise de dois conjuntos de dados exibidos no trabalho.

3.2 Métodos

A estimação dos parâmetros do modelo de regressão COM-Poisson de efeitos fixos é realizada maximizando uma forma reparametrizada da log-verossimilhança, definida na expressão 2.13, via algoritmo numérico de otimização *BFGS*. O parâmetro extra da COM-Poisson, ν tem suporte nos reais positivos, restringindo o espaço paramétrico de busca do otimizador, o que é numericamente indesejável. Para contornar essa empecilho reparametrizamos o modelo com o parâmetro $\phi = \log(\nu)$, assim como $0 < \nu < \infty$, $-\infty < \phi < \infty$. Sob a reparametrização a função a ser maximizada é

$$\ell(\phi, \beta \mid \underline{y}) = \sum_i^n y_i \log(\lambda_i) - e^\phi \sum_i^n \log(y_i!) - \sum_i^n \log(Z(\lambda_i, \phi)) \quad (3.1)$$

O ajuste do modelo é realizado sob ϕ . Portanto as inferências decorrentes do modelo são sobre esse parâmetro, todavia pode retornar para parametrização original utilizando a função inversa em valores pontuais ou método delta para funções de ϕ . Nesse trabalho as inferências são realizadas sob o parâmetro ϕ . Para esse parâmetro as interpretações são como se segue

$$\phi < 0 \Rightarrow \text{Superdispersão} \quad \phi = 0 \Rightarrow \text{Equidispersão} \quad \phi > 0 \Rightarrow \text{Subdispersão}$$

ou seja, possui a interpretação de um parâmetro de precisão.

Note que a partir dessa reparametrização a condução de testes de hipóteses é facilitada. Uma vez que $\phi = 0$, representa o caso particular em que a COM-Poisson se reduz a Poisson, a estatística

$$\begin{aligned} TRV &= 2 \cdot (\ell_{CMP} - \ell_P) \\ TRV &\sim \chi_1^2 \end{aligned}$$

em que ℓ_{CMP} e ℓ_P são as log-verossimilhanças maximizadas dos modelos COM-Poisson e Poisson com mesmo preditor linear respectivamente, se refere ao teste de razão de verossimilhanças para $H_0 : \phi = 0$, ou de forma mais apelativa, se refere ao teste sobre a equivalência dos modelos COM-Poisson e Poisson.

Para incluir um componente de barreira no modelo COM-Poisson, acomodando excesso de zeros, partimos da definição em 2.14 adotando para $Pr(Z = z \mid \Theta_c)$ a distribuição COM-Poisson (2.8) resultando em

$$Pr(Y = y \mid \pi, \phi, \lambda) = \begin{cases} \pi & \text{se } y = 0, \\ (1 - \pi) \frac{\lambda^y}{(y!)^{e^\phi} Z(\lambda, \phi)} \left(1 - \frac{1}{Z(\lambda, \phi)}\right)^{-1} & \text{se } y = 1, 2, \dots \end{cases} \quad (3.2)$$

Para modelos de regressão com componente de barreira, incluímos preditores lineares em π , $\underline{\pi} = \frac{\exp(G\gamma)}{1 + \exp(G\gamma)}$ e λ , $\underline{\lambda} = \exp(X\beta)$ e a verossimilhança desse modelo toma a forma

$$\mathcal{L}(\phi, \beta, \gamma \mid \underline{y}) = \mathbb{1}[\underline{\pi}] \cdot (1 - \mathbb{1}) \left[(1 - \underline{\pi}) \left(\frac{\underline{\lambda}^y}{(y!)^{e^\phi} Z(\underline{\lambda}, \phi)} \right) \left(1 - \frac{1}{Z(\underline{\lambda}, \phi)}\right) \right] \quad (3.3)$$

em que $\mathbb{1}$ é uma função indicadora para $y = 0$. Os argumentos $\hat{\phi}$, $\hat{\beta}$ e $\hat{\gamma}$, que maximizam o logaritmo neperiano da função 3.3 serão as estimativas de máxima verossimilhança do modelo COM-Poisson com componente de barreira.

Uma outra extensão proposta para o modelo COM-Poisson é a inclusão de efeitos aleatórios a fim de modelar a estrutura experimental ou observacional de um conjunto de dados. Nos reestringimos a inclusão de efeitos aleatórios Normais, ou seja,

$b \sim \text{Normal}(0, \Sigma)$, que são incorporados sob a forma $\underline{\lambda} = X\beta + Zb$ conforme especificação 2.16. Assim, considerando a distribuição COM-Poisson para a variável resposta condicionada as covariáveis e os efeitos aleatórios, podemos escrever a verossimilhança como

$$\mathcal{L}(\phi, \Sigma, \beta \mid \underline{y}) = \prod_{i=1}^m \int_{\mathbb{R}^q} \left(\prod_{j=1}^{n_i} \frac{\underline{\lambda}^y}{(y!)^{e^\phi} Z(\underline{\lambda}, \phi)} \right) \cdot (2\pi)^{q/2} |\Sigma| \exp \left(-\frac{1}{2} b^t \Sigma^{-1} b \right) db_i \quad (3.4)$$

sendo m o número de grupos que compartilham do mesmo efeito aleatório, q o número de efeitos aleatórios (intercepto aleatório, inclinação e intercepto aleatórios, etc.) e n_i o número de observações no i -ésimo grupo. A integração em 3.4, necessária para a avaliação da verossimilhança não tem forma analítica. Utilizamos a aproximação de Laplace da forma como apresentada em Ribeiro Jr et al. (2012, pág. 141) para aproximação dessa integral. A estimação dos parâmetros é realizada via maximização de $\log(\mathcal{L}(\phi, \Sigma, \beta \mid \underline{y}))$ com métodos numéricos de otimização. Ressalta-se que esse é um procedimento computacionalmente intensivo, pois a cada iteração do algoritmo de maximização, m aproximações de Laplace para integrais de dimensão q são realizadas. Ainda, quando considerada a distribuição COM-Poisson para a variável resposta condicionalmente independente, temos também o cálculo de n_m constantes normalizadoras $Z(\lambda, \phi)$ (2.9) para cada m grupo em cada iteração do algoritmo de otimização. Toda essa estrutura hierárquica com procedimentos computacionais realizados a cada estágio, são extremamente sensíveis a aspectos de programação/implementação.

Para comparação entre os modelos COM-Poisson e demais modelos listados no capítulo 2 utilizamos essencialmente o valor maximizado da log-verossimilhança. Critérios de informação de Akaike (AIC) e Bayesiano (BIC) definidos como

$$\text{AIC} = k - 2\ell(\Theta_k, \underline{y}) \quad \text{BIC} = k \log(n) - 2\ell(\Theta_k, \underline{y}) \quad (3.5)$$

sendo k o número de parâmetros, n o número de observações e $\ell(\Theta_k, \underline{y})$ a log-verossimilhança maximizada do modelo definido pelo conjunto Θ_k de parâmetros, também são utilizados. Nas análises comparamos também os níveis descritivos nos testes de razão de verossimilhanças entre modelos encaixados. Nos modelos de regressão de efeitos fixos os valores preditos pelos modelos COM-Poisson e demais alternativas pertinentes são contrastados graficamente com bandas de confiança.

Para maximização numérica das log-verossimilhanças dos modelos de regressão COM-Poisson e suas extensões utilizamos um método de otimização quasi-Newton bastante popular, denominado *BFGS*. As informações do vetor gradiente (derivadas de primeira e matriz hessiana (derivadas de segunda ordem) são obtidos numericamente via aproximação de diferenças finitas (NOCEDAL; WRIGHT, 1995).

4 Resultados e Discussão

Nesse capítulo são apresentados os modelos de regressão COM-Poisson ajustados aos dados apresentados na seção 3.1.1.

4.1 Análise de dados de capulhos de algodão sob efeito de desfolha

Diante da estrutura do experimento apresentada na seção 3.1.1.1 foram propostos, por Zeviani et al. (2014), cinco preditores crescentes em complexidade que testam aspectos interesses sobre os fatores experimentais envolvidos no experimento. Abaixo expressamos os cinco preditores considerados.

Preditor 1: $g(\mu) = \beta_0$

Preditor 2: $g(\mu) = \beta_0 + \beta_1 \text{def}$

Preditor 3: $g(\mu) = \beta_0 + \beta_1 \text{def} + \beta_2 \text{def}^2$

Preditor 4: $g(\mu) = \beta_0 + \beta_{1j} \text{def} + \beta_2 \text{def}^2$

Preditor 5: $g(\mu) = \beta_0 + \beta_{1j} \text{def} + \beta_{2j} \text{def}^2$

onde j varia nos níveis de estágio fenológico da planta (1: vegetativo, 2: botão floral, 3: florescimento, 4: maça, 5: capulho) e $g(\mu)$ uma função de ligação entre o componente sistemático e o componente aleatório do modelo. A proposta desses preditores foi realizada de forma aninhada a fim de facilitar a condução de testes de hipóteses. O modelo 1 contém apenas o intercepto, e é ajustado apenas como ponto de partida para verificar como modelos mais estruturados melhoram o ajuste. O modelo 2 apresenta apenas o efeito de desfolha de forma linear, o modelo 3 é o modelo 2 somado um efeito de segunda ordem. O modelo 4, apresenta o efeito de desfolha linear mudando de acordo com o estágio de crescimento, e por fim o modelo 5 diz que não somente o efeito de primeira ordem muda com o estágio de crescimento, mais também o efeito de segunda ordem.

A seguir ajustamos os modelos Poisson e COM-Poisson como alternativas paramétricas à análise de dados e como alternativa semi-paramétrica a estimação via quase-verossimilhança Poisson. Na tabela 3 apresentamos os resultados dos três modelos ajustados aos cinco preditores considerados. O modelo COM-Poisson apresentou melhor desempenho dentre todos os preditores considerados quando comparado ao Poisson, indicado pelas maiores log-verossimilhanças e menores AIC's.

As estimativas dos parâmetros extras ϕ e σ^2 dos modelos COM-Poisson e Quasi-Poisson respectivamente, também são apresentadas na 3 e indicam subdispersão ($\phi > 0$ e $\sigma^2 < 1$). Note que, mesmo quando não consideramos covariáveis, preditor 1, a

Tabela 3 – Medidas de ajuste para avaliação e comparação entre preditores e modelos ajustados

Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$		
Preditor 1	1	-279,93	561,87					
Preditor 2	2	-272,00	548,00	15,86	1	6,81E-05		
Preditor 3	3	-271,35	548,71	1,29	1	2,56E-01		
Preditor 4	7	-258,67	531,35	25,36	4	4,26E-05		
Preditor 5	11	-255,80	533,61	5,74	4	2,19E-01		
COM-Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$	$\hat{\phi}$	$P(> \chi^2)$
Preditor 1	2	-272,48	548,96				0,551	1,13E-04
Preditor 2	3	-257,46	520,93	30,03	1	4,25E-08	0,794	6,97E-08
Preditor 3	4	-256,09	520,18	2,75	1	9,73E-02	0,816	3,29E-08
Preditor 4	8	-220,20	456,40	71,78	4	9,54E-15	1,392	1,75E-18
Preditor 5	12	-208,25	440,50	23,90	4	8,38E-05	1,585	1,80E-22
Quase-Poisson	np	deviance	AIC	F	diff np	$P(> F)$	$\hat{\sigma}^2$	$P(> \chi^2)$
Preditor 1	1	75,51					0,567	3,66E-04
Preditor 2	2	59,65		34,21	1	4,17E-08	0,464	5,13E-07
Preditor 3	3	58,36		2,81	1	9,62E-02	0,460	3,66E-07
Preditor 4	7	33,00		22,77	4	5,89E-14	0,278	9,15E-16
Preditor 5	11	27,25		5,96	4	2,18E-04	0,241	3,57E-18

np, número de parâmetros, diff ℓ , diferença entre log-verossimilhanças, F, estatística F baseada nas quasi-deviances, diff np, diferença entre o np.

Fonte: Elaborado pelo autor, adaptação de Zeviani et al. (2014, Tabela 1).

hipótese de equidispersão foi rejeitada pelo modelos COM-Poisson e Quasi-Poisson. Isso se reflete nos níveis descritivos dos testes de razão de verossimilhanças realizados, em que o modelo Poisson, em discordância com os demais, não indicou significância do efeito quadrático por nível de desfolha, preditor 5, pois superestima a variabilidade do processo. Esses resultados estão de acordos com os apresentados por [Zeviani et al. \(2014\)](#), onde um modelo *Gamma-Count* foi ajustado, destaca-se a similaridade entre as medidas de ajuste dos modelos COM-Poisson e *Gamma-Count*. Os valores das log-verossimilhanças maximizadas nos dois modelos difere somente nas casas decimais, para todos os preditores.

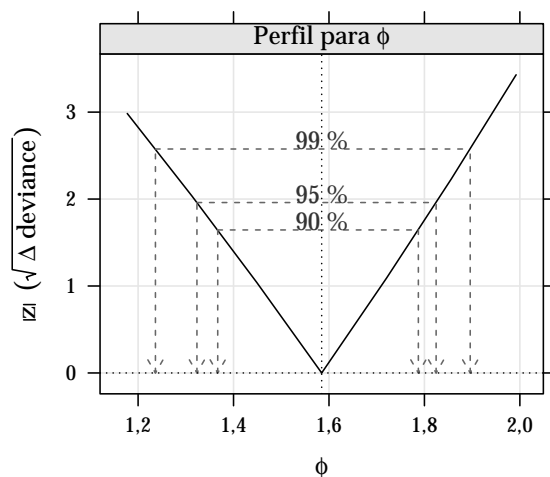
Na figura 17 apresentamos a avaliação do parâmetro ϕ do modelo COM-Poisson com efeito de desfolha artificial de primeira e segunda ordem para cada estágio fenológico, via verossimilhança perfilhada. O valor zero, que representa a não necessidade de um modelo COM-Poisson não é contemplado pelos intervalos de confiança de 90, 95 e até 99%. A simetria do perfil de verossimilhança também é algo para se destacar, pois neste caso intervalos do tipo Wald (computacionalmente mais fáceis), via aproximação quadrática da verossimilhança, podem ser construídos, muito embora os construídos via perfil de log-verossimilhança sejam preferíveis. Em concordância com a figura, o

teste de hipóteses via razão de verossimilhanças para $H_0 : \phi = 0$, rejeitou a hipótese nula com um nível de significância muito próximo a zero, tabela 3.

Tabela 4 – Estimativas dos parâmetros e razões entre as estimativa e erro padrão para os três modelos em estudo

Parâmetro	Poisson		Quasi-Poisson		COM-Poisson	
	Estimativa	Est/EP	Estimativa	Est/EP	Estimativa	Est/EP
ϕ	–	–	0,24	–	1,58	12,42
β_0	2,19	34,57	2,19	70,42	10,90	7,76
β_{11}	0,44	0,85	0,44	1,73	2,02	1,77
β_{12}	0,29	0,57	0,29	1,16	1,34	1,21
β_{13}	-1,24	-2,06	-1,24	-4,19	-5,75	-3,89
β_{14}	0,36	0,64	0,36	1,31	1,60	1,30
β_{15}	0,01	0,02	0,01	0,04	0,04	0,03
β_{21}	-0,81	-1,38	-0,81	-2,81	-3,72	-2,78
β_{22}	-0,49	-0,86	-0,49	-1,75	-2,26	-1,80
β_{23}	0,67	0,99	0,67	2,01	3,13	2,08
β_{24}	-1,31	-1,95	-1,31	-3,97	-5,89	-3,66
β_{25}	-0,02	-0,04	-0,02	-0,07	-0,09	-0,08

Fonte: Elaborado pelo autor.

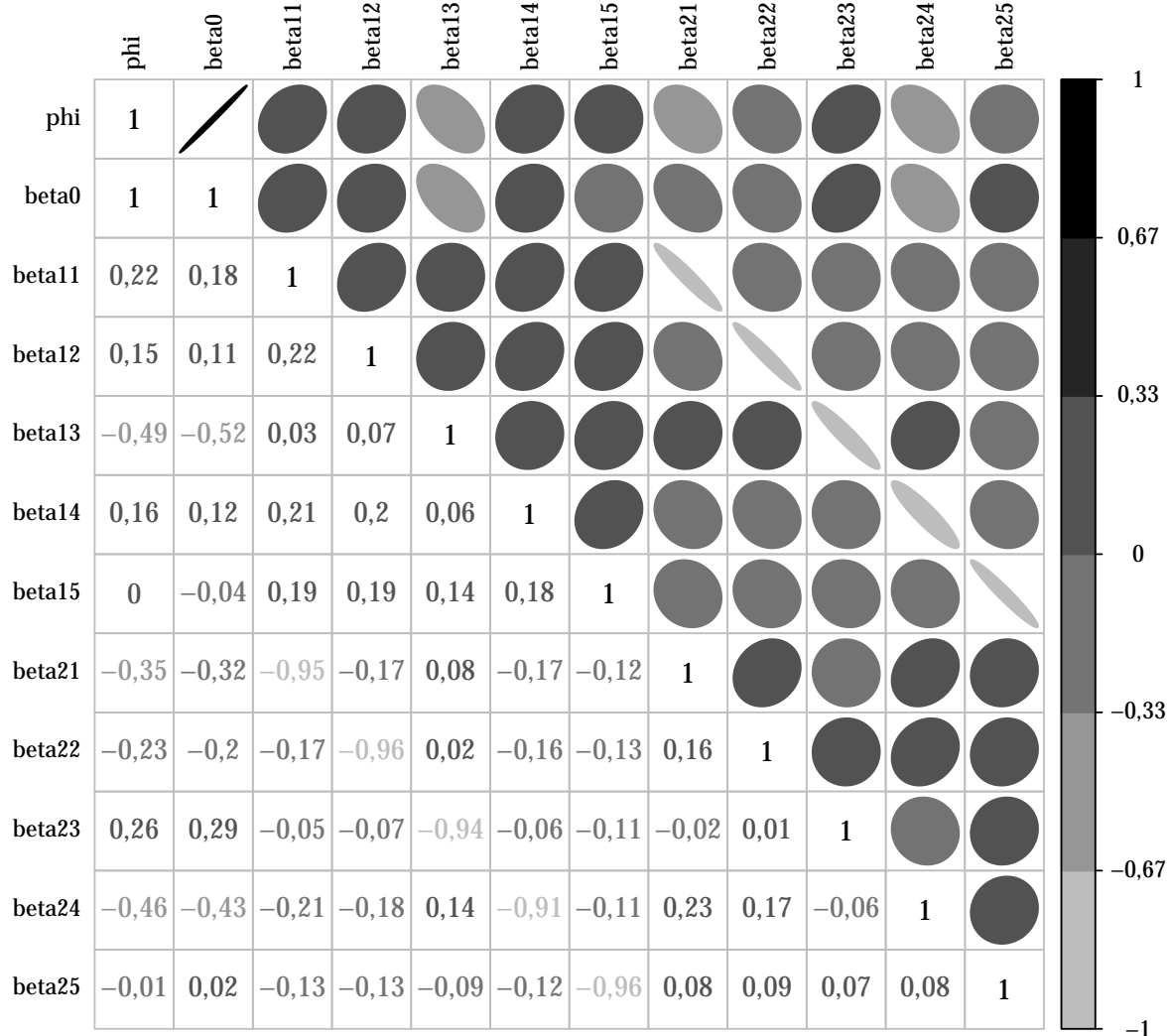


Fonte: Elaborado pelo autor.

Figura 17 – Perfil de log-verossimilhança para o parâmetro extra da COM-Poisson, estimado no modelo com o quinto preditor

As estimativas dos efeitos lineares e quadráticos de desfolha artificial, conforme notação do preditor 5, são apresentadas na tabela 4 para os modelos Poisson, Quasi-Poisson e COM-Poisson. As estimativas dos parâmetros para os modelos Poisson e Quasi-Poisson são idênticas, por construção 2.1, o que difere são as magnitudes dessas estimativas em comparação com seu erro padrão, que no caso Quasi-Poisson é corrigido pelo parâmetro σ^2 . Considerando o modelo COM-Poisson as estimativas são notavelmente diferentes, pois o preditor linear é construído em λ , da expressão 2.8, e este parâmetro não descreve, sozinho, a média da distribuição. Sendo assim as estimativas do COM-Poisson não podem ser comparadas com as demais estimativas. Contudo, a magnitude desses efeitos com relação ao efeito

padrão sim e neste caso temos os modelos Quasi-Poisson e COM-Poisson levando as mesmas conclusões.



Fonte: Elaborado pelo autor.

Figura 18 – Imagem da matriz de correlação entre os parâmetros do modelo COM-Poisson

As covariâncias entre as estimativas dos parâmetros do modelo COM-Poisson são apresentadas, na escala da correlação, na figura 18. Destaca-se nessa figura a forte correlação do parâmetro extra ϕ com os β 's da regressão. Embora seja uma representação empírica, observada a esse particular conjunto de dados, nota-se a não ortogonalidade na matriz de informação observada, o que implica que inferências sob o parâmetro ϕ influenciam as demais inferências a ser realizadas sobre os β 's. Esse comportamento dos modelos COM-Poisson é recorrente, como veremos também nos demais conjuntos de dados.

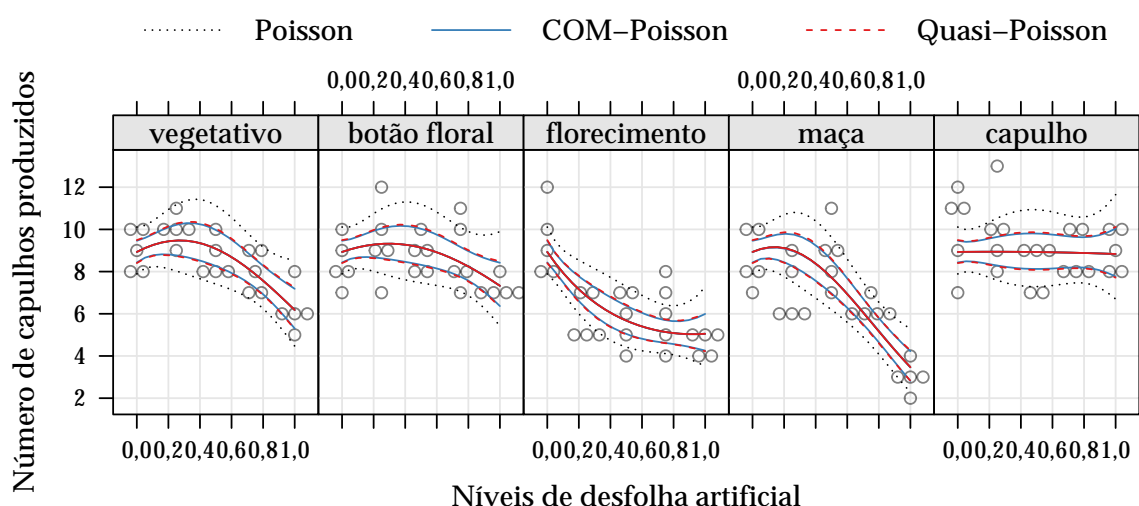


Figura 19 – Curva dos valores preditos com intervalo de confiança de (95%) como função do nível de desfolha e do estágio fenológico da planta.

Essa característica de não ortogonalidade da matriz de informação observada teve de ser levada em consideração para cálculo dos valores preditos, uma vez que a informação sobre a incerteza das estimativas contida na matriz de variâncias e covariâncias não pôde ser marginalizada para os β 's, que efetivamente são utilizados para cálculo de $\hat{\lambda}_i$ e consequentemente $\hat{\mu}_i$. Portanto, para cálculo dos valores preditos utilizamos a matriz de variâncias e covariâncias condicionada a ϕ (FERREIRA, 2011, teorema 3.6, pág. 123). Essa é uma prática tomada também para cálculo dos valores preditos nos demais conjunto de dados.

As médias com intervalos de confiança calculadas com os modelos COM-Poisson e Quasi-Poisson são praticamente idênticas, conforme pode ser visto na 19. Contudo, ressaltamos que o modelo COM-Poisson é totalmente paramétrico e assim conseguimos recuperar a distribuição, calculando probabilidades o que não é possível com a formulação Quasi-Poisson. Ainda nota-se claramente que o modelo Poisson é inadequado a esse conjunto de dados e que inferências a partir deste seriam incorretas.

4.2 Análise de dados de capulhos de algodão sob efeito de Mosca-Branca

Nesse conjunto de dados também temos indícios de subdispersão para as três variáveis de interesse mensuradas no estudo, conforme apresentado na seção 3.1.1.2. Para as três contagens procedeu-se com o ajuste dos modelos Poisson, Quasi-Poisson e COM-Poisson os preditores:

Preditor 1: $g(\mu) = \beta_0$

Preditor 2: $g(\mu) = \beta_0 + \beta_1 \text{dexp}$

Preditor 3: $g(\mu) = \beta_0 + \beta_1 \text{dexp} + \beta_2 \text{dexp}^2$

sendo dexp a variável aleatória dias de exposição à alta infestação de mosca-branca. Assim temos os preditores 1, 2, 3 que representam efeito nulo, linear e quadrático dos dias de exposição respectivamente.

Tabela 5 – Medidas de ajuste para avaliação e comparação entre preditores e modelos ajustados

np	Poisson			COM-Poisson			Quasi-Poisson	
	ℓ	AIC	$P(> \chi^2)$	ℓ	AIC	$P(> \chi^2)$	deviance	$P(> F)$
Número de capulhos produzidos								
1	-105,27	212,55		-92,05	188,09		20,80	
2	-105,03	214,05	4,83E-01	-91,31	188,62	2,25E-01	20,31	2,23E-01
3	-104,44	214,88	2,78E-01	-89,47	186,95	5,52E-02	19,13	6,16E-02
Número de estruturas reprodutivas								
1	-104,74	211,49		-86,41	176,82		16,23	
2	-104,27	212,54	3,32E-01	-84,59	175,18	5,66E-02	15,29	6,19E-02
3	-104,06	214,12	5,16E-01	-83,73	175,47	1,90E-01	14,87	2,07E-01
Número de nós da planta								
1	-143,79	289,59		-120,58	245,16		12,69	
2	-143,48	290,95	4,25E-01	-119,03	244,06	7,87E-02	12,05	7,39E-02
3	-142,95	291,89	3,04E-01	-116,27	240,54	1,88E-02	11,00	2,23E-02

np, número de parâmetros.

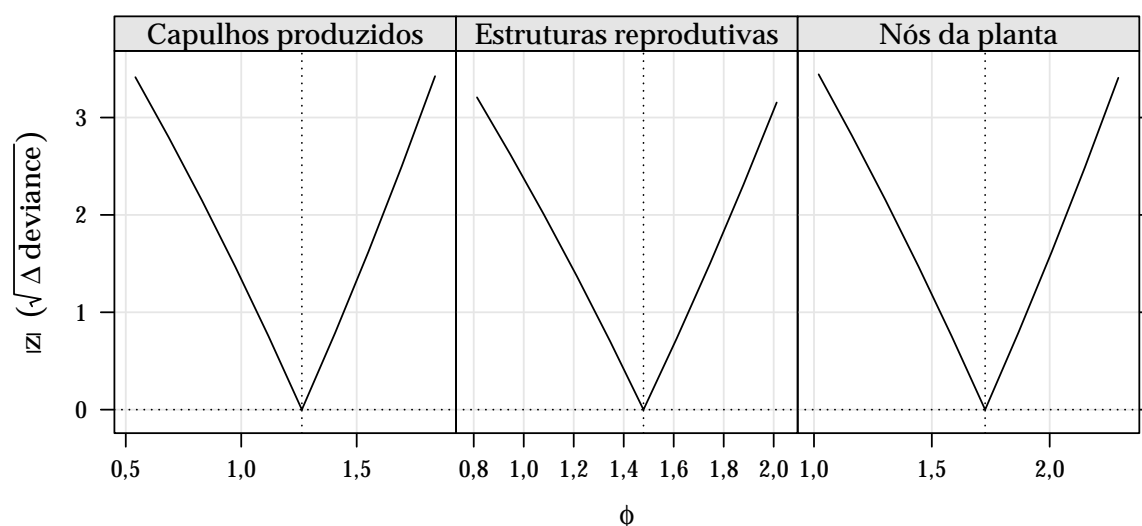
Fonte: Elaborado pelo autor.

Na tabela 5 são exibidos as medidas de ajuste dos modelos para as três variáveis resposta. Em todos os casos o modelo COM-Poisson apresentou maiores log-verossimilhanças indicando um melhor ajuste, quando comparado ao Poisson, também indicado pelos os valores de AIC que ponderam a log-verossimilhança pelo número de parâmetros considerados no modelo. Para questões inferenciais novamente, temos um desacordo entre os modelos paramétricos. Pelos modelos Poisson não temos evidências para manutenção de nenhum efeito da variável número de dias sob infestação, em todos os casos, ao passo que no modelo COM-Poisson indicou efeito quadrático quando considerado o modelo para o número de nós da planta (nível descritivo de 0,981) e o número de capulhos produzidos (nível descritivo de 0,945, na borda da região de significância, mas com uma diminuição do AIC em favor do efeito quadrático). Quando modelado o número de estruturas reprodutivas o modelo COM-Poisson também não indicou efeito quadrático, contudo o efeito linear de dexp pode ser discutido uma vez que a significância do TRV foi de 0,055 e o AIC apresentou um pequeno aumento com relação ao modelo nulo. Consideramos nas demais inferências os preditores com efei-

tos linear, para o número de estruturas reprodutivas e quadrático, para o número de capulhos produzidos e número de nós da planta.

A especificação do modelo via Quasi-Verossimilhança Poisson obteve níveis descritivos menos favoráveis a rejeição da hipótese nula que o modelo COM-Poisson. Contudo, para escolha de preditores as mesmas tendências apontadas pelo COM-Poisson foram seguidas.

Para avaliação do parâmetro ϕ da COM-Poisson nos três modelos considerados, temos os intervalos de confiança construídos sob perfilhamento da verossimilhança na figura 20. Note que para nenhum dos modelos os intervalos de confiança de 90, 95 e 99% de confiança contiveram o valor de $\phi = 0$. Os valores estimados dos parâmetros nos modelos para número de capulhos, número de estruturas reprodutivas e número de nós da planta foram de 1,263, 1,479, 1,726 respectivamente, indicando subdispersão.

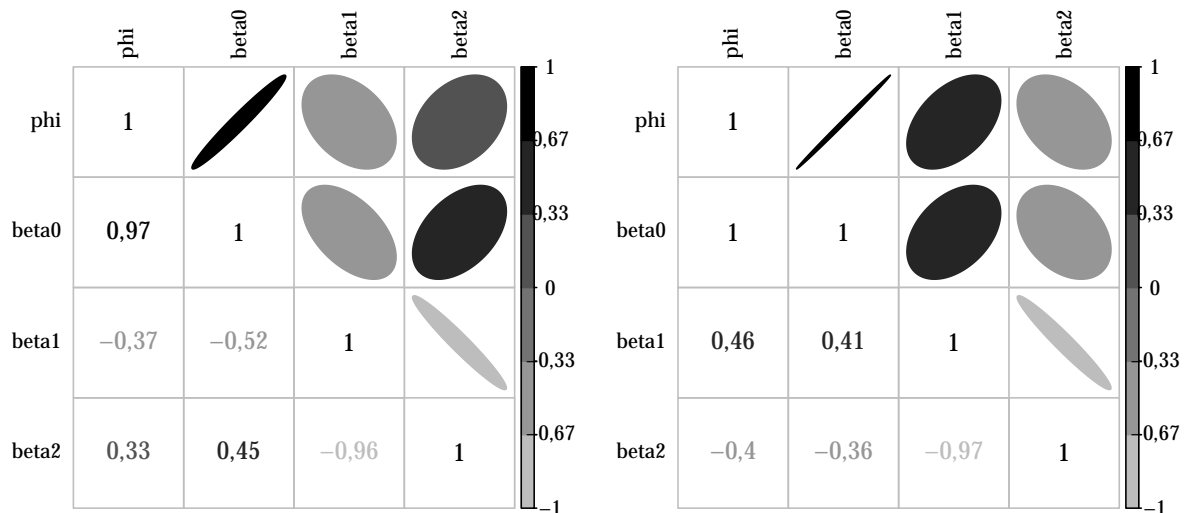


Fonte: Elaborado pelo autor.

Figura 20 – Perfis de log-verossimilhança para o parâmetro extra da COM-Poisson nos modelos para número de capulhos produzidos (esquerda), número de estruturas reprodutivas (central) e número de nós (direita).

Na figura 21 temos a representação da matriz de covariância entre as estimativas dos modelos para número de capulhos, à esquerda e número de nós da plantas, à direita. A forte correlação entre o parâmetro extra ϕ e β_0 (principalmente) também foi observada no ajuste do modelo para esses conjuntos de dados. No modelo considerado para o número de estruturas reprodutivas, o mesmo comportamento é observado, ??.

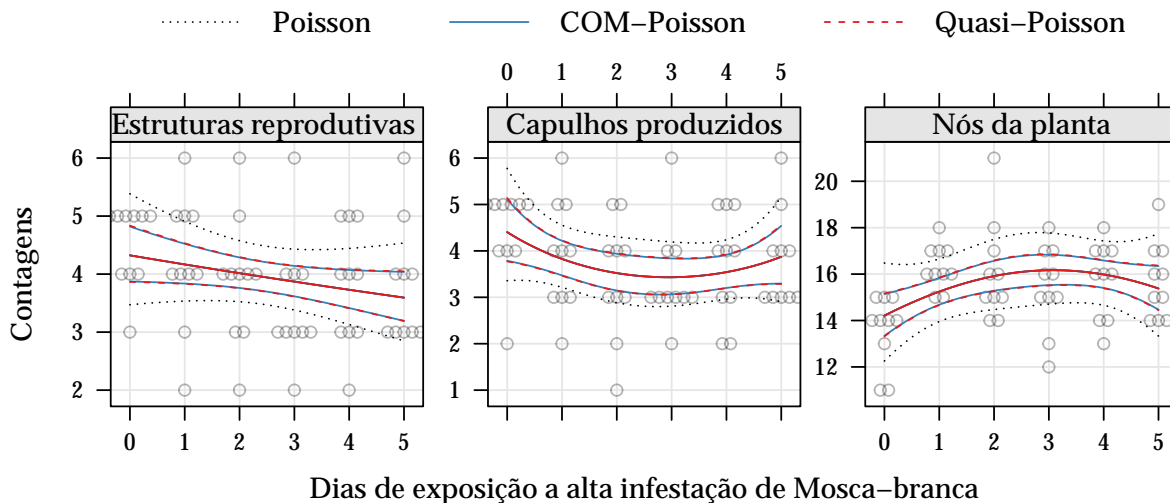
Finalmente calculando os valores preditos pelos modelos Poisson, COM-Poisson e Quasi-Poisson temos a representação gráfica na figura 23 com intervalos de confiança para média com 95% de confiança. Assim como na análise realizada na seção 4.1 temos os valores preditos com bandas de confiança obtidos dos modelos COM-Poisson e



Fonte: Elaborado pelo autor.

Figura 21 – Imagem da matriz de correlação entre os parâmetros do modelo COM-Poisson

Quasi-Poisson ajustados, praticamente idênticos levando as mesmas interpretações.

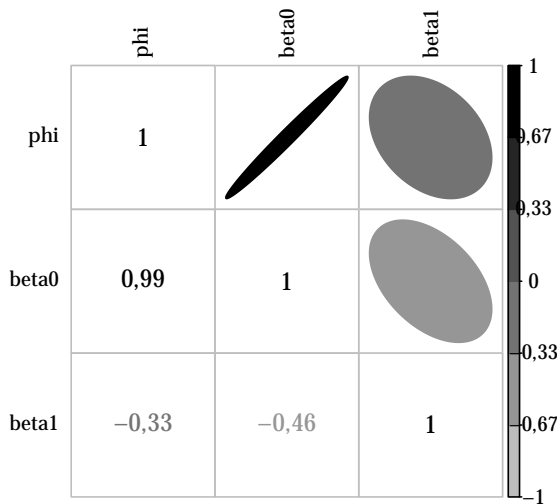


Fonte: Elaborado pelo autor.

Figura 23 – Curva dos valores preditos com intervalo de confiança de (95%) como função do nível de desfolha e do estágio fenológico da planta.

Com esse segundo exemplo de subdispersão, em que tivemos três contagens em um único experimento. Mostramos a flexibilidade do modelo COM-Poisson no que tange à característica de subdispersão, uma vez que seus resultados (predições pontuais e intervalares e testes de hipóteses para comparação de modelos) se equivalem a uma abordagem semi-paramétrica.

4.3 Análise de produção de soja sob efeito de umidade e adubação potássica



Fonte: Elaborado pelo autor.

Figura 22 – Imagem da matriz de correlação entre os parâmetros do modelo COM-Poisson

Nesse experimento apresentado em 3.1.1.3, também temos mais de uma variável de interesse em forma de contagem e pela descrição dos dados temos características relacionadas a dispersão da contagem distintas em ambas (equidispersão e superdispersão). Dos modelos apresentados no capítulo 2, o Poisson, COM-Poisson, Binomial-Negativo são as alternativas paramétricas avaliadas e o Quasi-Poisson é tomado como a alternativa semi-paramétrica. As variáveis de interesse números de grãos de soja e de vagens viáveis foram contabilizados por unidade experimental (vaso com duas plantas) e estão sob o efeito, controlado, de duas covariáveis, níveis de adubação potássica (0, 30, 60, 120, 180 mg dm⁻³) e níveis de umidade do solo (37.5, 50, 62.5 % do volume

total dos poros), que consideramos na análise como fatores com 5 e 3 níveis respectivamente. Ainda têm-se pela condução do experimento o efeito relacionado a blocagem realizada, foram cinco blocos utilizados para controle de variação local. Os preditores considerados são

$$\text{Preditor 1: } \eta_1 = g(\mu_{ijk}) = \beta_0 + \tau_i + \gamma_j + \delta_k$$

$$\text{Preditor 1: } \eta_2 = g(\mu_{ijk}) = \beta_0 + \tau_i + \gamma_j + \delta_k + \alpha_{jk}$$

em que τ_i é o efeito do i -ésimo bloco, $i = 1$: bloco II, 2: bloco III, 3: bloco IV e 4: V; γ_j o efeito do j -ésimo nível de umidade aplicado, $j = 1$: 50% e 2: 62,5%; δ_k o efeito do k -ésimo nível de adubação potássica, $k = 30, 60, 120$ e 180 mg dm^{-3} e α_{jk} o efeito da interação entre o j -ésimo nível de umidade do solo e o k -ésimo nível de adubação potássica. Assim no modelo mais completo, com interação, teremos 19 parâmetros de locação a serem estimados.

Para ajuste dos modelos COM-Poisson nesse exemplo tivemos um tempo computacional ligeiramente mais demorado (em torno de 10s para os quatro modelos considerando as duas contagens e os dois preditores). Isso se deve ao fato das contagens serem altas (variando entre 92 e 271 para o número de grãos e 36 e 110 para o número

de vagens) e estarmos em um caso superdisperso ($\phi < 0$). Nesse cenário a constante normalizada $Z(\lambda_i, \nu = \exp(\phi))$, expressão 2.9, converge para 0 mais lentamente.

Medidas de qualidade de ajuste calculadas sob os modelos Poisson, COM-Poisson, Binomial Negativo e Quasi-Poisson são apresentadas na tabela 6. Considerando a variável resposta número de vagens viáveis, percebe-se que há indícios de equidispersão indicados i) pelos parâmetros extras dos modelos alternativos ao Poisson, em que estimativas $\hat{\phi}$, $\hat{\theta}$ e $\hat{\sigma}^2$ estão próximas dos valores 0, ∞ e 1 que compreendem o caso particular Poisson nos modelos COM-Poisson, Binomial Negativo e Quasi-Poisson, ii) pelas log-verossimilhanças dos modelos paramétricos que resultaram em valores muito próximos, iii) pelos valores de AIC que foram menores nos modelos Poisson, mostrando que não há ganho expressivo quando estimados os parâmetros extra dos modelos alternativos. Os *p-valores* associados ao TRV entre os modelos COM-Poisson e Poisson com preditores 1 e 2 foram de 0,671, 0,446, evidenciando a equidispersão dos dados. Na figura 24 à esquerda apresentamos os intervalos de confiança baseados no perfil de verossimilhança para ϕ no modelo COM-Poisson com efeito de interação, como esses intervalos exibidos contemplam o valor 0 o modelo COM-Poisson pode ser reduzido ao Poisson. Para avaliação dos preditores, temos novamente um caso de valores na borda de significância. Seguimos as análises permanecendo com o modelo mais completo que considera a interação entre adubação e umidade.

Tabela 6 – Medidas de ajuste para avaliação e comparação entre preditores e modelos ajustados

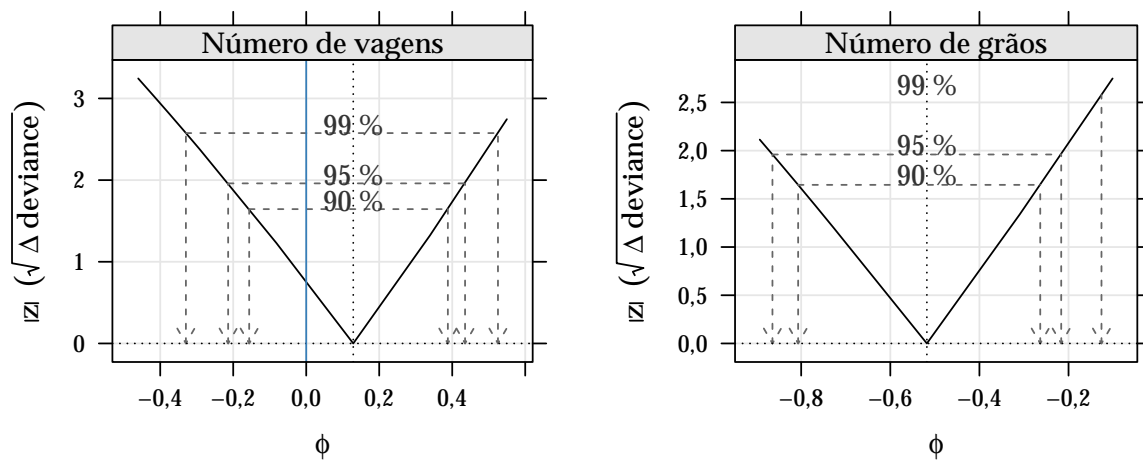
PO	np	Número de vagens				Número de grãos			
		ℓ	AIC	$P(> \chi^2)$		ℓ	AIC	$P(> \chi^2)$	
η_1	11	-266,69	555,38			-343,16	708,33		
η_2	19	-259,62	557,23	7,79E-02		-321,67	681,34	8,83E-07	
CP	np	ℓ	AIC	$P(> \chi^2)$	$\hat{\phi}$	ℓ	AIC	$P(> \chi^2)$	$\hat{\phi}$
η_1	12	-266,60	557,20		-6,75E-02	-326,61	677,21		-8,17E-01
η_2	20	-259,33	558,65	6,85E-02	1,29E-01	-315,64	671,29	5,06E-03	-5,18E-01
BN	np	ℓ	AIC	$P(> \chi^2)$	$\hat{\theta}$	ℓ	AIC	$P(> \chi^2)$	$\hat{\theta}$
η_1	12	-266,69	557,37		4,59E+03	-326,54	677,07		1,42E+02
η_2	20	-259,62	559,23	7,82E-02	1,03E+06	-315,39	670,77	4,39E-03	2,61E+02
QP	np	ℓ	AIC	$P(> \chi^2)$	$\hat{\sigma}^2$	ℓ	AIC	$P(> \chi^2)$	$\hat{\sigma}^2$
η_1	11	79,43			1,28E+00	167,71			2,71E+00
η_2	19	65,28		1,87E-01	1,20E+00	124,72		3,00E-02	2,29E+00

np, número de parâmetros, PO, Poisson, CP, COM-Poisson, BN, Binomial Negativo, QP, Quasi-Poisson.

Fonte: Elaborado pelo autor.

No fragmento direito da tabela 6 são apresentados os resultados para os modelos

que ajustam o número de grãos por parcela. Neste caso temos evidências de superdispersão, pois as estimativas dos parâmetros ϕ e σ^2 foram menores que zero e maiores que 1 respectivamente. Os valores de AIC se apresentam menores e as avaliações da log-verossimilhança no ponto máximo maiores para os modelos paramétricos alternativos ao Poisson. Ainda a evidência sobre o efeito de interação para essa variável resposta é mais contundente. Na 24 à direita temos a verossimilhança perfilhada com indicação dos intervalos de confiança para ϕ e estes não contemplam o zero.



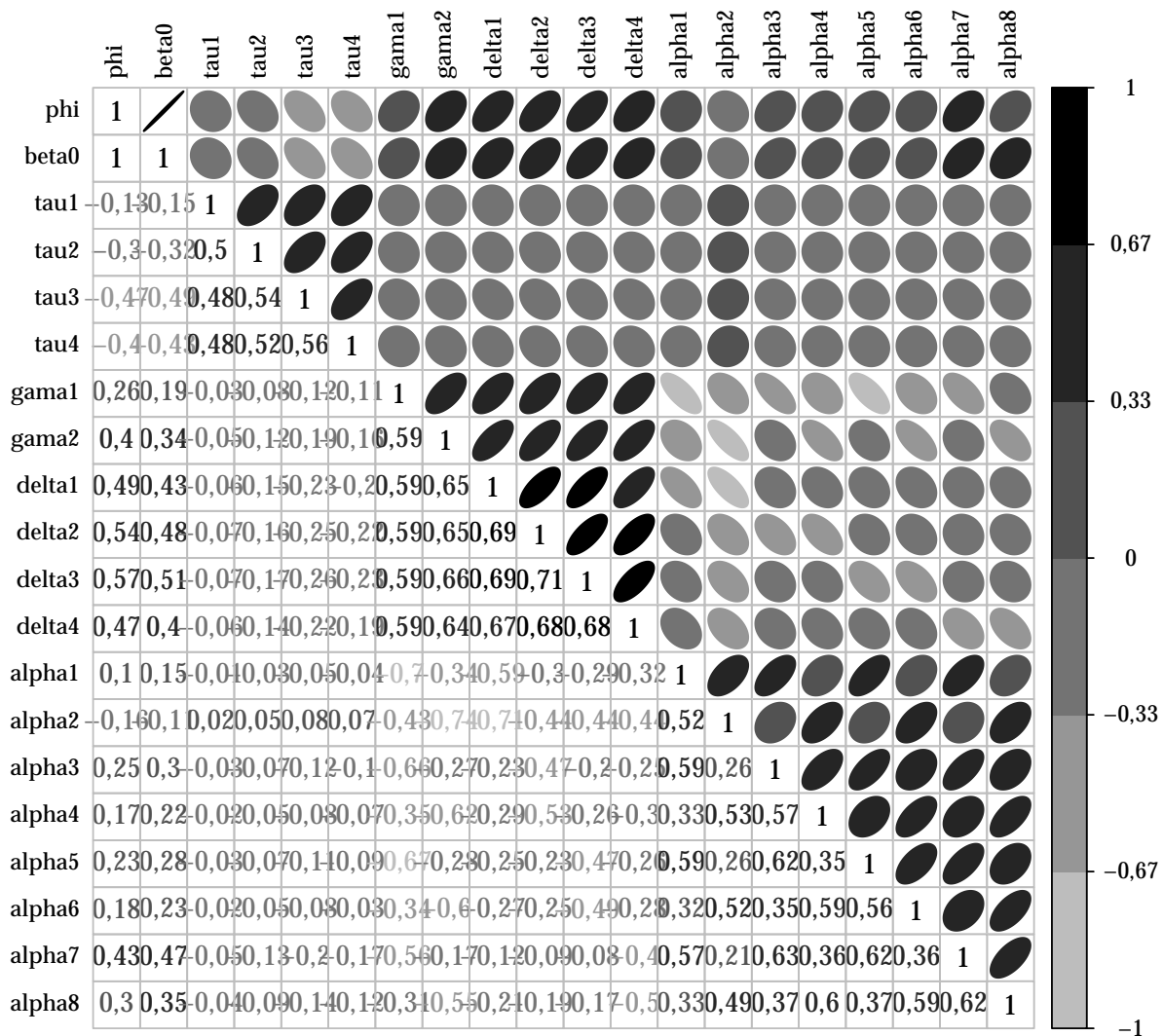
Fonte: Elaborado pelo autor.

Figura 24 – Perfis de log-verossimilhança para o parâmetro extra da COM-Poisson nos modelos para número de vagens viáveis por parcela (esquerda) e número grãos de soja por parcela (direita).

A visualização da covariâncias entre as estimativas dos parâmetros no modelo COM-Poisson para o número de vagens por parcela é feita na figura ?? e para o número de grãos por parcela na figura 25. Em ambos os casos a correlação entre os parâmetros de locação (β 's) e dispersão (ϕ) ganha destaque.

Na figura 27 apresentamos as médias calculadas com com intervalos de confiança 95% sob os modelos Poisson, COM-Poisson, Binomial-Negativo e Quasi-Poisson considerando efeito de interação entre os níveis de umidade do solo e adubação potássica de confiança. Tomou-se o efeito médio de bloco, uma vez que esse efeito não é de interesse prático.

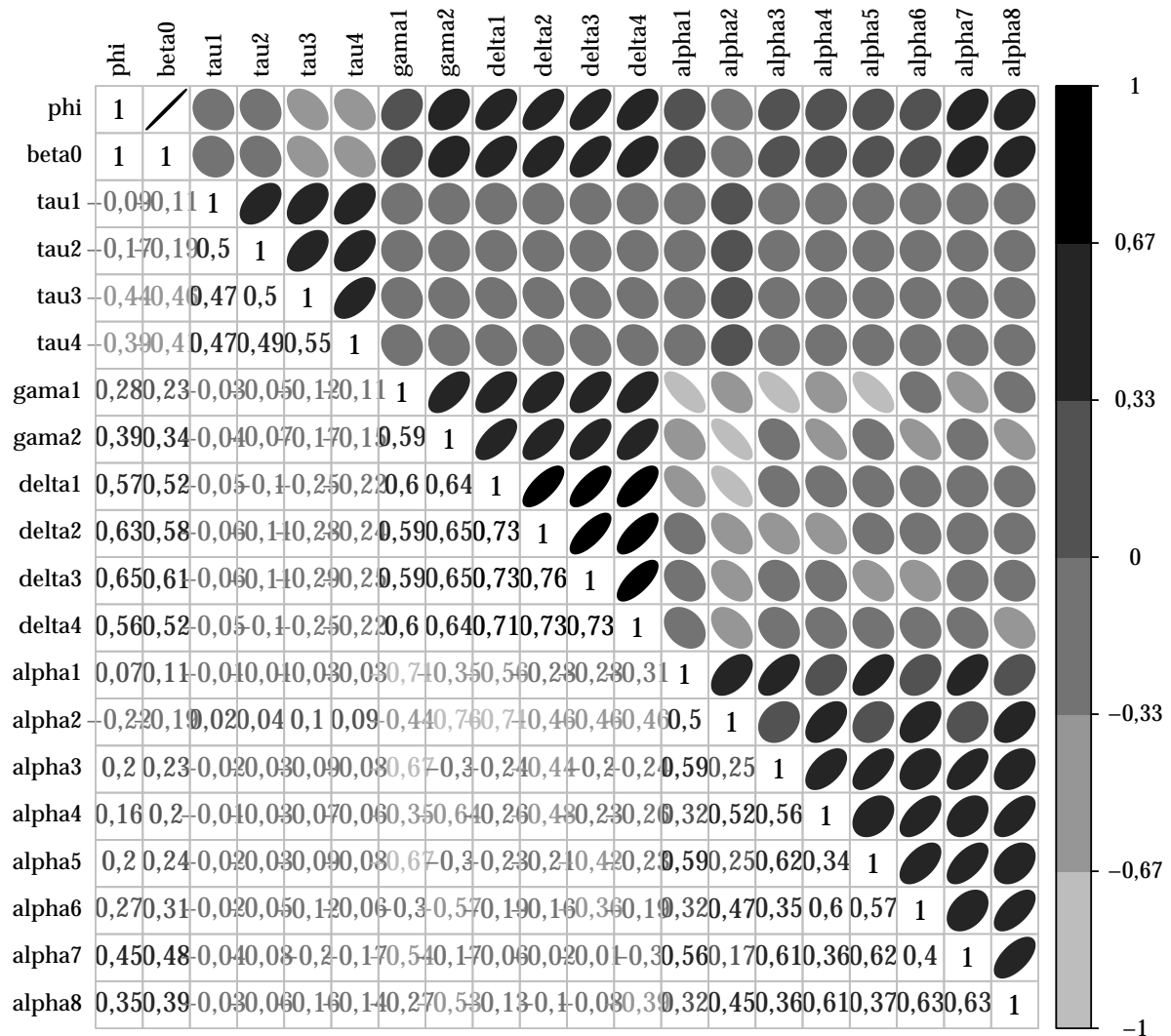
Para a contagem do número de vagens, observa-se os intervalos com comprimento muito parecidos, ligeiramente menores para o caso COM-Poisson e Binomial Negativo. Para a contagem do número de grão por parcela, onde temos um caso superdisperso, percebe-se que o modelo Poisson nos leva a falsa confiança, uma vez que os intervalos são menores não por se ajustar melhor aos dados, mas sim por subesti-



Fonte: Elaborado pelo autor.

Figura 25 – Imagem da matriz de correlação entre os parâmetros do modelo COM-Poisson ajustados ao número de vagens por parcela.

mar a variabilidade do processo. Para as formulações alternativas, temos os modelos paramétricos com intervalos menores que o semi-paramétrico Quasi-Poisson, isso é razoável, pois nos Quasi-Poisson temos somente a especificação de dois momentos, enquanto que nos paramétricos especificamos a distribuição completa, ganhando informação 2.4. Os intervalos sob os modelos COM-Poisson e Binomial Negativa foram os mais parcimoniosos, sendo intervalos menores, porém fiéis a variabilidade inerente ao processo.

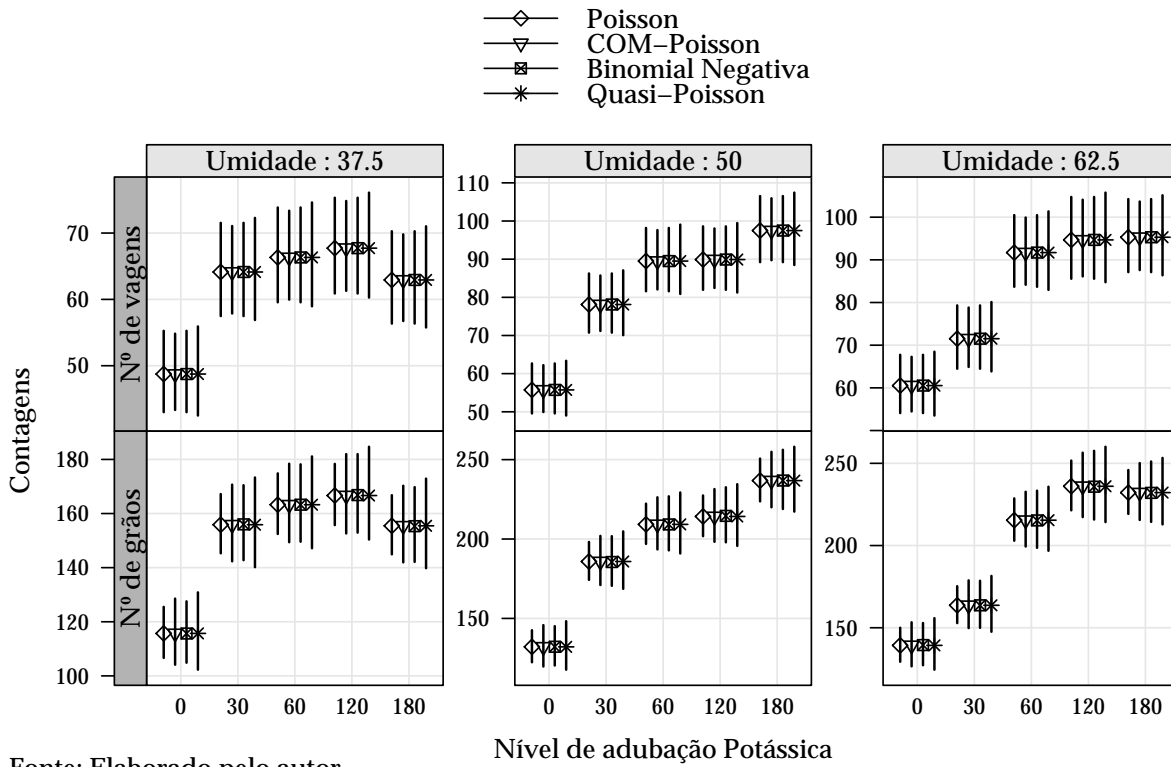


Fonte: Elaborado pelo autor.

Figura 26 – Imagem da matriz de correlação entre os parâmetros do modelo COM-Poisson ajustados ao número de grãos por parcela.

4.4 Análise de ninfas de mosca-branca em lavoura de soja

Neste experimento também temos fortes indícios de superdispersão, conforme visto na seção 3.1.1.4. Assim os modelos Poisson, COM-Poisson, Binomial Negativo e Quasi-Poisson serão aplicados. A variável em estudo é a contagem de ninfas de Mosca-Branca nos folíolos de plantas de soja e têm-se interesse na avaliação dos fatores dias decorridos após a primeira avaliação da planta e cultivar. Como o experimento foi conduzido sob delineamento de blocos casualizados, os efeitos de bloco são considerados no modelo. As covariáveis serão tratadas como fator, assim como na aplicação anterior, com seis níveis para o número de dias decorridos a partir da primeira avaliação e quatro nível para o fator cultivar de soja. Os preditores em comparação são:



Fonte: Elaborado pelo autor.

Figura 27 – Valores preditos com intervalos de confiança (95%) como função do nível de adubação com potássio e do percentual de umidade do solo para cada variável de interesse mensurada (número de vagens e número de grãos por parcela).

Preditor 1: $\eta_1 = g(\mu_{ijk}) = \beta_0 + \tau_i + \gamma_j + \delta_k$

Preditor 1: $\eta_2 = g(\mu_{ijk}) = \beta_0 + \tau_i + \gamma_j + \delta_k + \alpha_{jk}$

em que τ_i é o efeito do i -ésimo bloco, $i = 1$: bloco II, 2: bloco III, 3: bloco IV e 4: V; γ_j o efeito da j -ésima cultivar, $j = 1$: BRS 243 RR, 2: BRS 245 RR e 3: BRS 246 RR; δ_k o efeito do k -ésimo nível do número de dias após o início do experimento, $k = 8, 13, 22, 31$ e 38 dias e α_{jk} o efeito da interação entre a j -ésima cultivar e o k -ésimo nível do número de dias após o início do experimento. Como construído, a avaliação do efeito de interação é de interesse prático, pois informa se há desempenhos distintos das cultivares conforme o número de dias decorridos da primeira avaliação. No modelo com interação, temos 27 parâmetros de locação a serem estimados.

Assim como na aplicação superdispersa apresentada na seção 4.3, nesse exemplo temos um cenário com contagens altas (variando entre 92 e 271) e ainda superdispersas (parâmetros ϕ estimados próximos à -3). Isso torna a convergência da função $Z(\lambda_i, \nu = \exp(\phi))$ demorada e o valor dessa constante que normaliza a densidade é altíssimo. Em problemas com contagens altas e comportamento muito superdisperso a obtenção da constante Z pode se tornar proibitiva computacionalmente, devido à *over-*

flow (valores que ultrapassam o limite de capacidade de armazenamento da máquina) e consequentemente o modelo COM-Poisson não se ajusta.

Nesse exemplo, os modelos COM-Poisson convergiram e seus resultados são exibidos na tabela 7 em conjunto com os resultados proporcionados pelos modelos Poisson, Binomial Negativo e COM-Poisson. Todas as estimativas dos parâmetros extras nos modelos concorrentes ao Poisson, $\hat{\phi}$, $\hat{\theta}$ e $\hat{\sigma}^2$ indicam expressivamente a superdispersão os dados. Em benefício dos modelos alternativos ao Poisson temos todas as medidas apresentadas indicando uma substancial melhora de ajuste quando flexibilizamos o modelo. Destaque para a magnitude dessas evidências, em que, por exemplo, o AIC obtido dos modelos alternativos é em torno de 0,47 vezes o obtido do Poisson.

Tabela 7 – Medidas de ajuste para avaliação e comparação entre preditores e modelos ajustados

Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$	
Preditor 1	12	-922,98	1869,96				
Preditor 2	27	-879,23	1812,46	87,50	15	2,90E-12	
COM-Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$	$\hat{\phi}$
Preditor 1	13	-410,44	846,89				-3,08
Preditor 2	28	-407,15	870,30	6,59	15	9,68E-01	-2,95
Binomial Neg.	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$	$\hat{\theta}$
Preditor 1	13	-406,16	838,31				3,44
Preditor 2	28	-400,55	857,10	11,21	15	7,38E-01	3,99
Quase-Poisson	np	deviance	AIC	F	diff np	$P(>F)$	$\hat{\sigma}^2$
Preditor 1	12	1371,32					17,03
Preditor 2	27	1283,82		0,31	15	9,93E-01	19,03

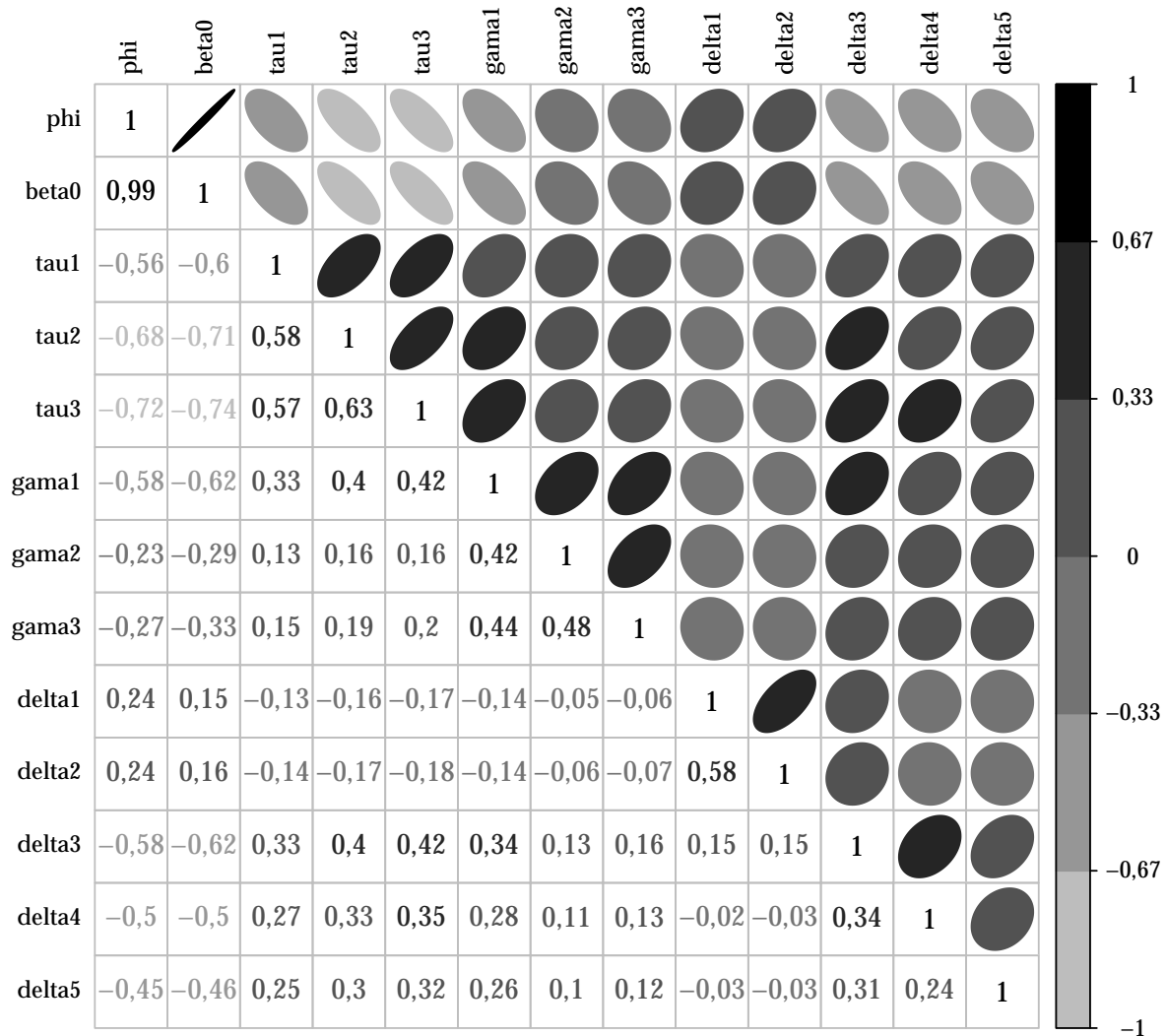
np, número de parâmetros, diff ℓ , diferença entre log-verossimilhanças, F, estatística F baseada nas quasi-deviances, diff np, diferença entre o np.

Fonte: Elaborado pelo autor.

Para tomada de decisão, observa-se que o modelo Poisson é claramente inadequado. Para avaliação dos preditores, na tabela 7, temos o modelo Poisson indicando (com uma significância inferior a 1E-10) que há efeito de interação entre os dias decorridos da primeira avaliação e as cultivares ao passo que nos modelos alternativos esse efeito é marcadamente não significativo. Essa discordância se deve, conforme já discutido, ao fato de o modelo Poisson subestimar a variabilidade por sua reestrição de equidispersão. Assim, com variâncias menores qualquer efeito acrescido no modelo passará por significativo.

Enfatizando a superdispersão indicada pelo modelo COM-Poisson, considerando o preditor de efeitos aditivos, temos o perfil de verossimilhança para o parâmetro

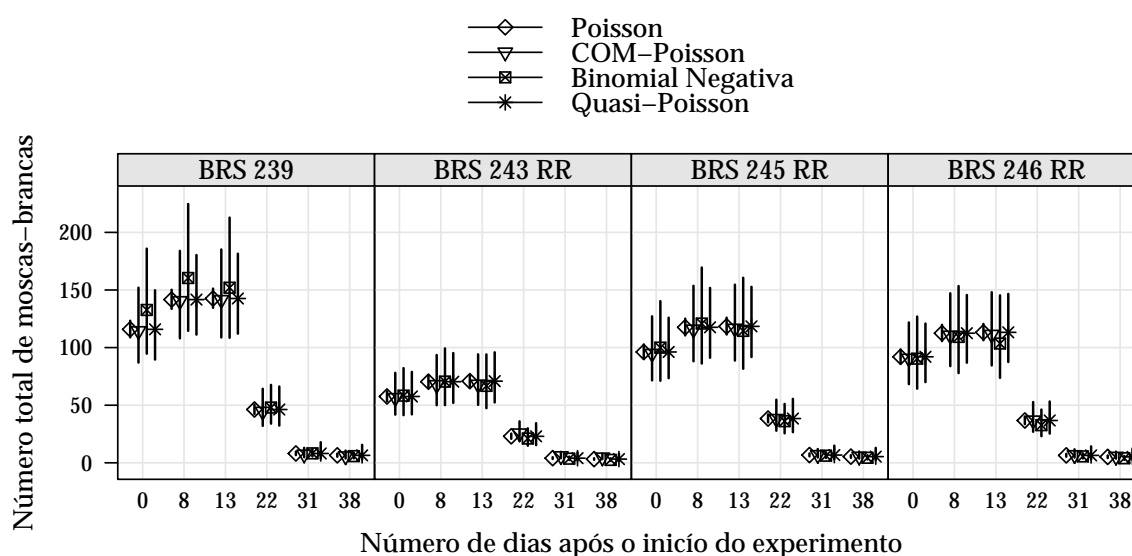
ϕ apresentado na figura 28. Podemos observar que os limites inferiores dos intervalos de confiança de 90, 95 e 99% estão muito distantes do valor 0, sob o qual temos equivalência entre os modelos Poisson e COM-Poisson. Outra característica desse gráfico é a leve assimetria à esquerda, o que atribuímos a forte característica de superdispersão dos dados.



Fonte: Elaborado pelo autor.

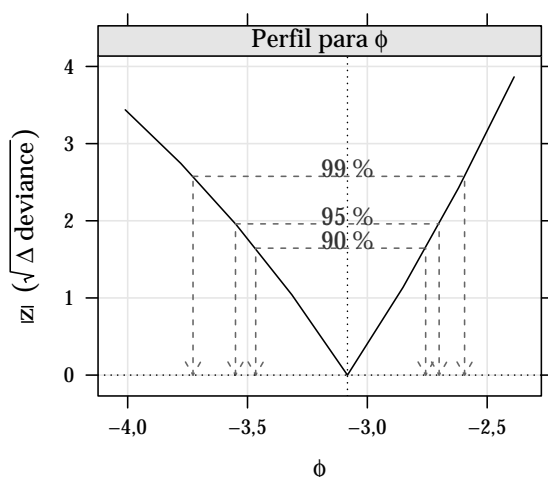
Figura 29 – Imagem da matriz de correlação entre os parâmetros do modelo COM-Poisson

Também exibimos as covariâncias entre os efeitos estimados pelo modelo COM-Poisson conforme descrição do preditor 1 na figura 29 na escala de correlação. Similarmente as análises anteriores observa-se a alta correlação entre $\hat{\phi}$ e os demais parâmetros de regressão. A soma dos valores absolutos das correlações observadas entre $\hat{\phi}$ e as demais estimativas é de 7,059 e a média 0,543.



Fonte: Elaborado pelo autor.

Figura 30 – Valores preditos com intervalos de confiança (95%) em função das cultivares de soja e da data de avaliação da planta.



Fonte: Elaborado pelo autor.

Figura 28 – Perfil de log-verossimilhança para o parâmetro extra da COM-Poisson

As médias com intervalos de confiança calculadas para cada combinação dos níveis de dias após a primeira avaliação e cultivar de soja considerando os modelos Poisson, COM-Poisson, Binomial-Negativo e Quasi-Poisson, são apresentadas na 30. Para o efeito de bloco consideramos o efeito médio para uma correta comparação. Podemos observar que o intervalo de confiança descrito pelo modelo Poisson é quase imperceptível quando comparados aos demais, mostrando novamente que seu uso é inadequado a esses dados. Já para as outras alternativas não tivemos um comportamento padrão em todas as cultivares. Os intervalos pelo modelo Quasi-Poisson e COM-Poisson foram muito similares em todos os casos

e os intervalos pelo modelo Binomial Negativo intervalos mais amplos. Um fato interessante é que não necessariamente as estimativas pontuais da média desses modelos alternativos serão iguais, isso ocorre, por construção, somente para nos modelos Poisson e Quasi-Poisson, esse exemplo ilustra na prática a constatação desse fato. Para o mo-

delo Binomial Negativo tivemos médias visivelmente superiores que os demais para a cultivar BRS 239. Para o modelo COM-Poisson as estimativas pontuais são visivelmente iguais as do modelo Poisson.

4.5 Análise de captura de peixes em um parque estadual

Nesse exemplo ilustramos a análise de um estudo observacional em que aparentemente temos uma quantidade excessiva de contagens nulas (veja a seção 3.1.1.5). O estudo tem por objetivo a modelagem do número de peixes capturados por grupos de visitantes em um Parque Estadual. As covariáveis mensuradas foram np, o número de pessoas no grupo, nc, o número de crianças e ca variável binária que indica a presença ou não de um campista no grupo.

Como já antecipado pela visualização e apresentação dos dados, modelos estruturados de forma convencional, que pressupõe apenas um processo estocástico na geração de dados, não se ajustaram adequadamente. A seguir apresentamos a alternativa de inclusão de um efeito de barreira para acomodar a quantidade excessiva de valores zero. Os modelos Poisson, Binomial Negativo e COM-Poisson sob esta estruturação são ajustados e comparados.

Com a estrutura dos dados vamos modelar o número de peixes capturados em duas partes, as contagens nulas e as não nulas, seção 2.4. Abaixo definimos os preditores considerados para as duas partes

$$\begin{aligned} \text{Preditor 1: } g(\mu) &= \beta_0 + \beta_1 ca + \beta_2 np \\ \text{logit}(\pi) &= \gamma_0 + \gamma_1 ca + \gamma_2 np + \gamma_3 nc \end{aligned}$$

$$\begin{aligned} \text{Preditor 2: } g(\mu) &= \beta_0 + \beta_1 ca + \beta_2 np + \beta_3 nc + \beta_4(np \cdot nc) \\ \text{logit}(\pi) &= \gamma_0 + \gamma_1 ca + \gamma_2 np + \gamma_3 nc + \gamma_4(np \cdot nc) \end{aligned}$$

sendo $g(\mu)$ e $\text{logit}(\pi)$ as funções de ligação que relacionam os preditores lineares com as médias dos modelos para contagens não nulas e contagens zero respectivamente. Os preditores lineares foram propostos de forma aninhada. No primeiro temos os efeitos aditivos de todas as covariáveis mensuradas para a parte das contagens nulas e efeitos aditivos do número de pessoas e de crianças para a parte das contagens não nulas. No segundo temos os efeitos aditivos de todas as covariáveis acrescido do efeito de interação entre o número de pessoas e de crianças para ambas as partes do modelo.

Na tabela 8 as medidas de ajuste dos modelos Poisson, Binomial Negativo e COM-Poisson são apresentadas para comparação dos resultados. Observa-se pelas log-verossimilhanças maximizadas que o modelo Poisson não se ajustou adequadamente quando comparado aos demais. Isso se deve ao fato discutido na seção 2.4, que mesmo

Tabela 8 – Medidas de ajuste para avaliação e comparação de preditores e modelos com componente de barreira ajustados

Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	P(> χ^2)	
Preditor 1	7	-857,48	1728,96				
Preditor 2	10	-744,58	1509,17	225,79	3	1,12E-48	
Binomial Negativo	np	ℓ	AIC	2(diff ℓ)	diff np	P(> χ^2)	$\hat{\theta}$
Preditor 1	8	-399,79	815,58				0,20
Preditor 2	11	-393,72	809,44	12,14	3	6,91E-03	0,37
COM-Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	P(> χ^2)	$\hat{\phi}$
Preditor 1	8	-409,85	835,71				-8,77
Preditor 2	11	-402,30	826,59	15,12	3	1,72E-03	-3,77

np, número de parâmetros, diff ℓ , diferença entre log-verossimilhanças, F, estatística F baseada nas quasi-devianças, diff np, diferença entre o np.

Fonte: Elaborado pelo autor.

modelando os zeros podemos ter diferentes níveis de dispersão para as contagens nulas. Nesse exemplo as contagens não nulas são superdispersas, conforme visto pelas estimativas dos parâmetros extras do modelo Binomial Negativo e COM-Poisson. Conforme indicado pelos níveis descritivos dos TRV's aplicados nos modelos encaixados temos a indicação de que o modelo com efeitos de interação é distinto do modelo com efeitos aditivos definidos no preditor 1.

As estimativas dos parâmetros para cada especificação de modelos são exibidas na tabela 9. Observe primeiramente que as estimativas dos parâmetros $\gamma_i, i = 0, 1, 2, 3, 4$ são idênticas, independentemente do modelo adotado. Esse resultado é esperado, pois na construção dos modelos com componente de barreira, a modelagem da parte que contempla os valores zero é realizada via distribuição Bernoulli com parâmetro $\pi = \text{logit}(Z\gamma)$. As diferenças entre os modelos ocorre na distribuição considerada para a parte das contagens não nulas.

Nos efeitos estimados para a parte da modelagem dos valores não nulos têm-se algumas diferenças consideráveis. Destaca-se que o valor das estimativas dos modelos Poisson e Binomial Negativo são comparáveis entre si, pois modelam a média da distribuição, mas não comparáveis com as estimativas do modelo COM-Poisson, pois este modelo um parâmetro que não representa, unicamente, a média. Contudo, independente da distribuição o sinal dos efeitos deve ser o mesmo. Isso não ocorre nas estimativas dos parâmetros β_3 , positiva no modelo Poisson e negativa nos demais e β_4 , positiva no modelo COM-Poisson e negativa nos demais. Porém esses efeitos não tem impacto significativo para definição dos parâmetros das distribuições, conforme podemos ver na figura ?? que mostra as médias calculadas com base nas três formulações. A seguir discutimos sobre os valores apresentados dos erros padrão dessas estimativas.

Tabela 9 – Estimativas dos parâmetros e razões entre as estimativa e erro padrão para os três modelos em estudo

Parâmetro	Poisson		Binomial Negativo		COM-Poisson	
	Estimativa	Est/EP	Estimativa	Est/EP	Estimativa	Est/EP
Extra (ϕ, θ)			0,37	-2,08	-3,77	-9,52
β_0	-1,01	-5,44	-1,75	-2,90	-0,62	-29,74
β_1	0,74	7,88	0,41	1,23	0,10	29,20
β_2	0,89	18,55	1,05	6,41	0,14	21,86
β_3	0,49	1,11	-0,06	-0,05	-0,33	-17,53
β_4	-0,45	-3,69	-0,32	-0,90	0,04	33,41
γ_0	-2,58	-5,08	-2,58	-5,08	-2,59	-5,09
γ_1	0,98	3,00	0,98	3,00	1,00	3,04
γ_2	1,25	5,60	1,25	5,60	1,26	5,61
γ_3	-0,93	-1,05	-0,93	-1,05	-0,93	-1,06
γ_4	-0,41	-1,41	-0,41	-1,41	-0,41	-1,41

Fonte: Elaborado pelo autor.

Calculando a magnitude desses efeitos quando escalonados pelo seu erro padrão, calculado pelo negativo do inverso da matriz hessiana, temos diferenças substanciais. O modelo COM-Poisson indica erros padrões das estimativas muito menores que os apresentados no modelo Binomial Negativo. Sob investigações do problema, encontramos que este resultado se deve por inconsistências no procedimento numérico para determinação da matriz hessiana por diferenças finitas no modelo COM-Poisson. Portanto, os erros padrão sob o modelo COM-Poisson apresentados não são confiáveis.

Embora tenhamos constatado problemas nos algoritmos numéricos para determinar a curvatura da log-verossimilhança, temos que as estimativas pontuais são coerentes com os demais modelos, conforme visto no figura 31 onde temos apresentadas as médias calculadas com base nos três modelos estudados. Observa-se em todos os modelos a mesma tendência.

Com esse ilustramos a extensão do modelo COM-Poisson para acomodar excesso de zeros e ressaltamos que nesse exemplo tivemos contagens não nulas superdispersas e para esses casos temos a distribuição Binomial Negativa sendo a principal alternativa. Porém em casos que as contagens não nulas se apresentam de forma subdispersa não temos opções prontamente disponíveis para análise e o modelo COM-Poisson com componente de barreira, conforme apresentado, se torna uma abordagem atrativa.

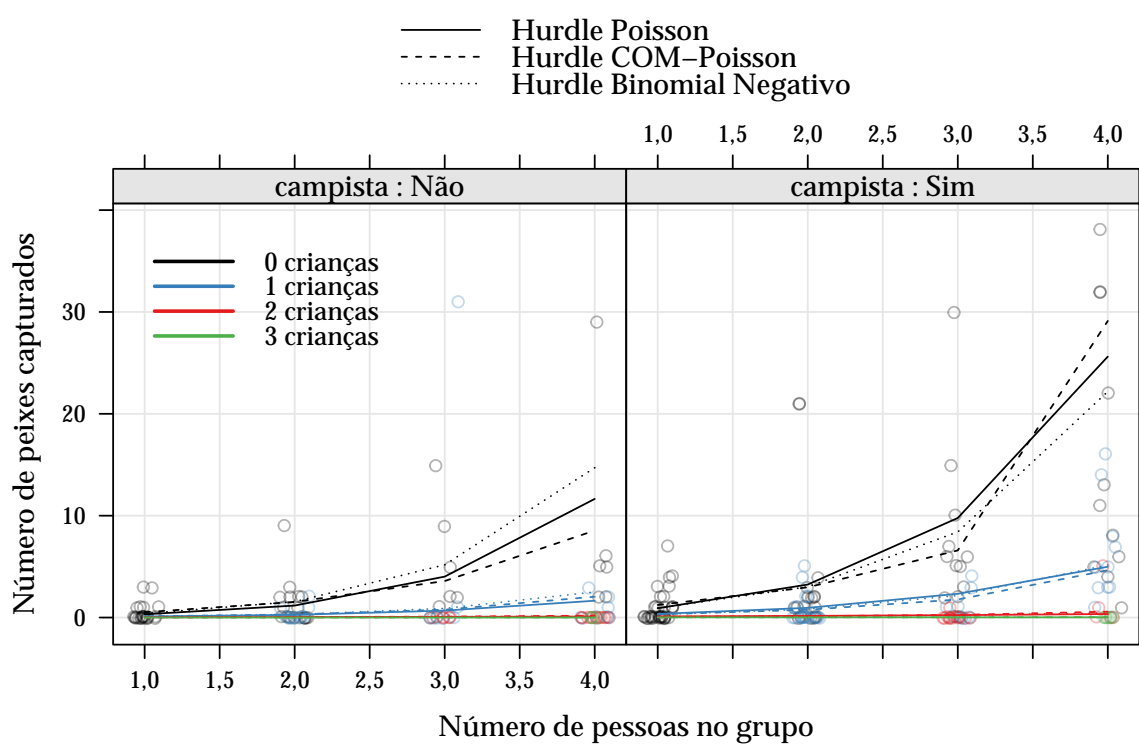


Figura 31 – Valores preditos do número de peixes capturados considerando o número de crianças e pessoas no grupo e a presença de um campista

5 Considerações Finais

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

REFERÊNCIAS

- BATES, D. M. et al. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, v. 67, p. 1–48, 2015. ISSN 0092-8615. Disponível em: <http://lme4.r-forge.r-project.org/IMMwR/lrgprt.pdf>. Citado na página 34.
- BORGES, P. *Novos modelos de sobrevivência com fração de cura baseados no processo da carcinogênese*. Tese (Doutorado) — Universidade Federal de São Carlos, 2012. Citado na página 20.
- CONWAY, R. W.; MAXWELL, W. L. A queuing model with state dependent service rates. *Journal of Industrial Engineering*, v. 12, p. 132—136, 1962. Citado 2 vezes nas páginas 19 e 27.
- FERREIRA, D. F. *Estatística Multivariada*. [S.l.]: Editora UFLA, 2011. Citado na página 49.
- HILBE, J. M. *Modeling Count Data*. [S.l.: s.n.], 2014. 300 p. ISSN 1467-9280. ISBN 978-1-107-02833-3. Citado 2 vezes nas páginas 21 e 32.
- KING, G. Variance specification in event count models: from restrictive assumptions to a generalized estimator. *American Journal of Political Science*, v. 33, n. 3, p. 762–784, aug 1989. ISSN 00925853. Disponível em: <http://www.jstor.org/stable/2111071>. Citado na página 17.
- KOKONENDJI, C. C. Over- and Underdispersion Models. In: *Methods and Applications of Statistics in Clinical Trials: Planning, Analysis, and Inferential Methods*. [s.n.], 2014. p. 506–526. Disponível em: https://lmb.univ-fcomte.fr/IMG/pdf/ch30{_}kokonendji2014. Citado na página 21.
- LAMBERT, D. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*, v. 34, n. 1, p. 1, feb 1992. ISSN 00401706. Disponível em: <http://www.jstor.org/stable/1269547?origin=crossref>. Citado 2 vezes nas páginas 22 e 32.
- LORD, D.; GEEDIPALLY, S. R.; GUIKEMA, S. D. Extension of the application of conway-maxwell-poisson models: Analyzing traffic crash data exhibiting underdispersion. *Risk Analysis*, v. 30, n. 8, p. 1268–1276, 2010. ISSN 02724332. Citado na página 21.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, v. 135, p. 370–384, 1972. Citado 2 vezes nas páginas 17 e 24.
- NOCEDAL, J.; WRIGHT, S. J. *Numerical optimization*. Springer, 1995. 636 p. ISSN 0011-4235. ISBN 0387987932. Disponível em: <https://books.google.com/books?id=epc5fX0lqRIC&pgi>. Citado na página 44.
- PAULA, G. A. *Modelos de regressão com apoio computacional*. IME-USP São Paulo, 2013. Disponível em: <https://www.ime.usp.br/~giapaula/textoregressao.h>. Citado 2 vezes nas páginas 18 e 25.

- RIBEIRO, A. M. T. *Distribuição COM-Poisson na análise de dados de experimentos de quimioprevenção do câncer em animais*. Tese (Doutorado) — Universidade Federal de São Carlos, 2012. Citado 2 vezes nas páginas 20 e 33.
- Ribeiro Jr, P. J. et al. Métodos computacionais para inferência com aplicações em R. In: *20º Simpósio Nacional de Probabilidade e Estatística*. [s.n.], 2012. p. 282. Disponível em: <<http://leg.ufpr.br/doku.php/cursos:mcie>>. Citado 2 vezes nas páginas 18 e 44.
- RIDOUT, M.; DEMETRIO, C. G.; HINDE, J. Models for count data with many zeros. *International Biometric Conference*, n. December, p. 1–13, 1998. Citado 2 vezes nas páginas 22 e 32.
- SELLERS, K. F.; RAIM, A. A flexible zero-inflated model to address data dispersion. *Computational Statistics & Data Analysis*, Elsevier B.V., v. 99, p. 68–80, jul 2016. ISSN 01679473. Disponível em: <<http://dx.doi.org/10.1016/j.csda.2016.01.007><http://linkinghub.elsevier.com/retrieve/pii/S0167947316000165>>. Citado na página 20.
- SELLERS, K. F.; SHMUELI, G. A flexible regression model for count data. *Annals of Applied Statistics*, v. 4, n. 2, p. 943–961, 2010. ISSN 19326157. Citado 3 vezes nas páginas 21, 27 e 29.
- SERAFIM, M. E. et al. Umidade do solo e doses de potássio na cultura da soja. *Revista Ciência Agronômica*, v. 43, n. 2, p. 222–227, jun 2012. ISSN 1806-6690. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-66902012000200003&lng=pt&nrm>. Citado na página 38.
- SHMUELI, G. et al. A useful distribution for fitting discrete data: Revival of the Conway-Maxwell-Poisson distribution. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, v. 54, n. 1, p. 127–142, 2005. ISSN 00359254. Citado 3 vezes nas páginas 19, 28 e 29.
- SILVA, A. M. et al. Impacto de diferentes níveis de desfolha artificial nos estádios fenológicos do algodoeiro. *Revista de Ciências Agrárias*, v. 35, n. 1, p. 163–172, 2012. ISSN 0871-018X. Disponível em: <<http://www.cabdirect.org/abstracts/20123299470.html;jsessionid=CF06663390A3A4463413D3018ECAACD6>>. Citado na página 35.
- SUEKANE, R. *DISTRIBUIÇÃO ESPACIAL E DANO DE MOSCA-BRANCA Bemisia tabaci (GENNADIUS, 1889) BIÓTIPO B NA SOJA*. Tese (Doutorado) — Universidade Federal da Grande Dourados, 2011. Citado na página 39.
- WEDDERBURN, R. W. M. Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika*, v. 61, n. 3, p. 439, dec 1974. ISSN 00063444. Disponível em: <<http://www.jstor.org/stable/2334725?origin=crossref>>. Citado na página 24.
- WINKELMANN, R. Duration Dependence and Dispersion in Count-Data Models. *Journal of Business & Economic Statistics*, v. 13, n. 4, p. 467–474, oct 1995. ISSN 0735-0015. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/07350015.1995.10524620>>. Citado 2 vezes nas páginas 18 e 23.
- WINKELMANN, R. *Econometric Analysis of Count Data*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. 342 p. ISBN 978-3-540-77648-2. Disponível

em: <<http://medcontent.metapress.com/index/A65RM03P4874243N.pdf><http://link.springer.com/10.1007/978-3-540-78389-3>>. Citado na página 21.

WINKELMANN, R.; ZIMMERMANN, K. F. *Count data models for demographic data*. 1994. 205–221, 223 p. Citado na página 18.

ZEILEIS, A.; KLEIBER, C.; JACKMAN, S. Regression Models for Count Data in R. *Journal Of Statistical Software*, v. 27, n. 8, p. 1076–84, 2007. ISSN 1465735X. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21518631>>. Citado na página 32.

ZEVIANI, W. M. et al. The Gamma-count distribution in the analysis of experimental underdispersed data. *Journal of Applied Statistics*, n. October, p. 1–11, 2014. ISSN 0266-4763. Disponível em: <<http://dx.doi.org/10.1080/02664763.2014.922168>>. Citado 5 vezes nas páginas 21, 23, 35, 45 e 46.

Apêndices

APÊNDICE A – Lipsum

Nulla ac nisl. Nullam urna nulla, ullamcorper in, interdum sit amet, gravida ut, risus. Aenean ac enim. In luctus. Phasellus eu quam vitae turpis viverra pellentesque. Duis feugiat felis ut enim. Phasellus pharetra, sem id porttitor sodales, magna nunc aliquet nibh, nec blandit nisl mauris at pede. Suspendisse risus risus, lobortis eget, semper at, imperdiet sit amet, quam. Quisque scelerisque dapibus nibh. Nam enim. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc ut metus. Ut metus justo, auctor at, ultrices eu, sagittis ut, purus. Aliquam aliquam.

Anexos

ANEXO A – Lipsum

Sed mattis, erat sit amet gravida malesuada, elit augue egestas diam, tempus scelerisque nunc nisl vitae libero. Sed consequat feugiat massa. Nunc porta, eros in eleifend varius, erat leo rutrum dui, non convallis lectus orci ut nibh. Sed lorem massa, nonummy quis, egestas id, condimentum at, nisl. Maecenas at nibh. Aliquam et augue at nunc pellentesque ullamcorper. Duis nisl nibh, laoreet suscipit, convallis ut, rutrum id, enim. Phasellus odio. Nulla nulla elit, molestie non, scelerisque at, vestibulum eu, nulla. Ut odio nisl, facilisis id, mollis et, scelerisque nec, enim. Aenean sem leo, pellentesque sit amet, scelerisque sit amet, vehicula pellentesque, sapien.