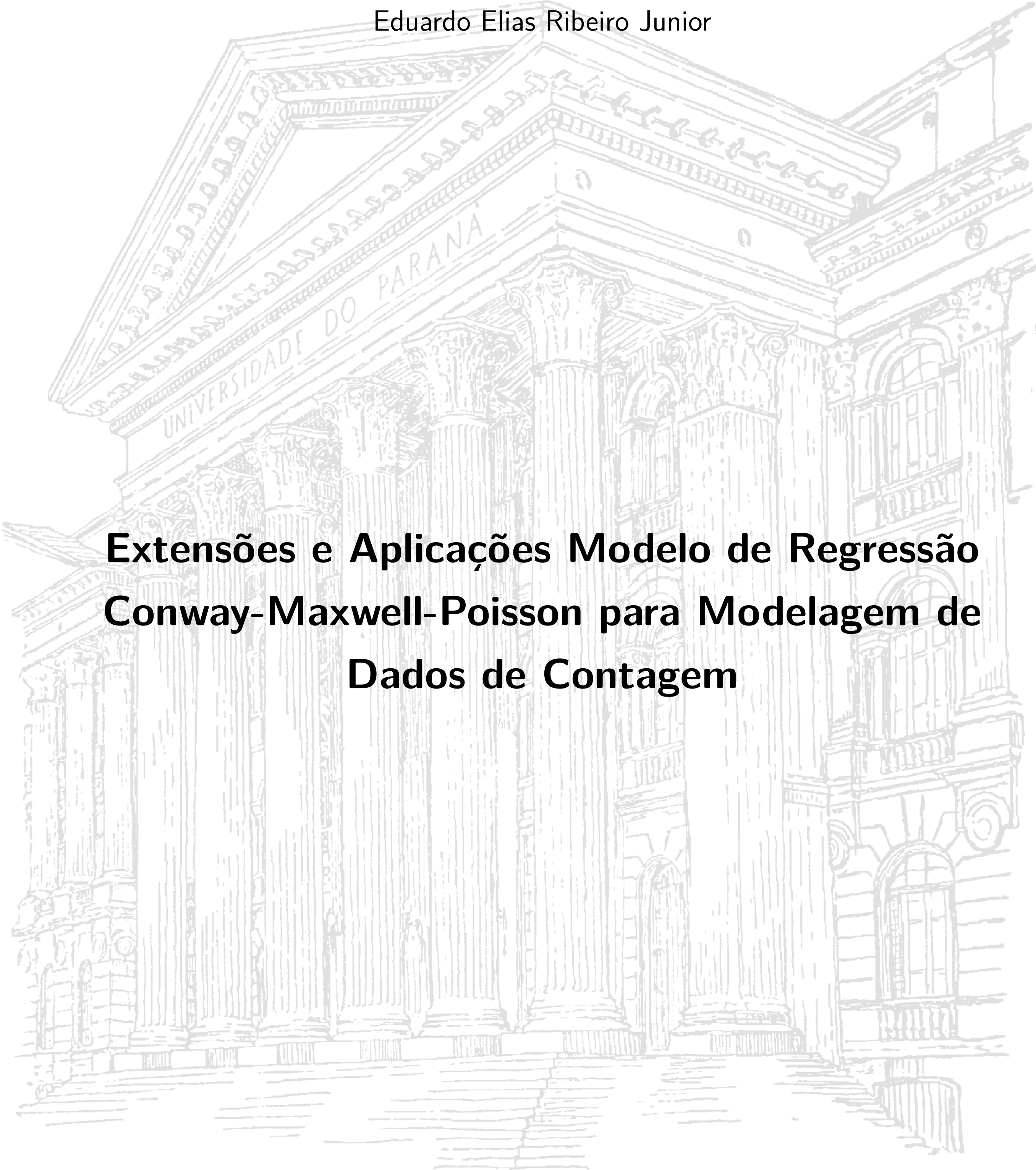
Universidade Federal do Paraná

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, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer 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.

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

Resumo

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer 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

Lista de tabelas

Lista de símbolos

log	Logarítmo neperiano (de base e)

Sumário

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 predições 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 uma 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 me 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 predições 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 intepretação 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 parâmentros 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

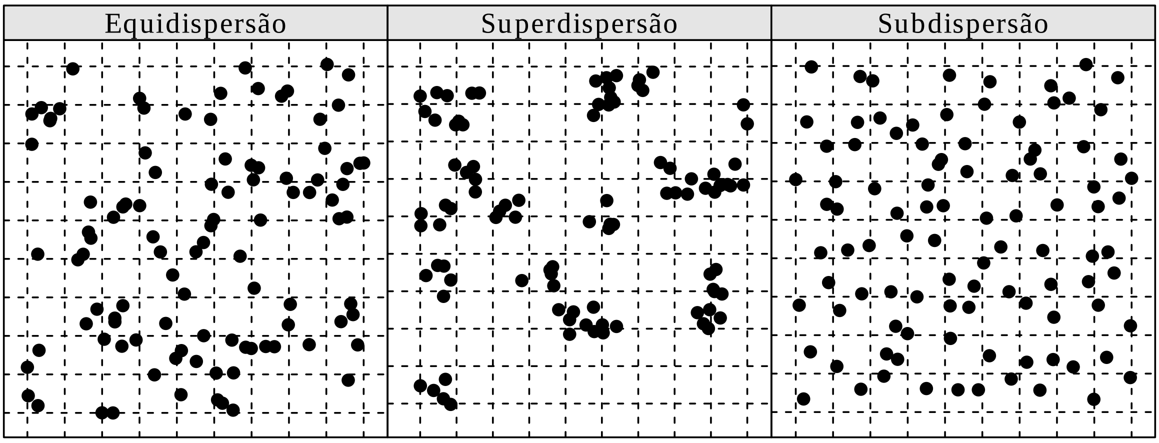


Figura 1 – Ilustração de diferentes tipos de processos pontuais. Da direita para esquerda temos processos sob padrões aleatório, aglomerado e uniforme

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 co contagens mutio 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 dispõe 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 alterativa 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 em homenagem aos seus autores Richard W. Conway, William L. Maxwell, **Co**nway-**M**axwell-Poisson) generaliza a distribuição Poisson com a adição de mais uma parâmetro, denotado por ν, que torna a razão de probabilidades sussecivas 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 Sobrevivência, 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 estudas.

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 **??** apresentammos os conjuntos de dados a serem analisados e os métodos para ajuste e comparação dos modelos. O capítulo 4 traz os 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 equidipersão, 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	✓	✓	
Inverse Gaussian Poisson	✓	✓	
Compound Poisson	✓	✓	
Poisson Generalizada	✓	✓	✓
Gamma-Count	✓	✓	✓
COM-Poisson	✓	✓	✓
Katz	✓	✓	✓
Poisson Polynomial	✓	✓	✓
Double-Poisson	✓	✓	✓
Lagrangian Poisson	✓	✓	✓

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 perimitem 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

$$Pr(Y = y \mid \lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \qquad y = 0, 1, 2, \cdots \tag{2.1}$$

em que $\lambda > 0$ representa a taxa de ocorrência do evento de interesse. Uma particularidade já destacada desta distribuição é que $E(X) = V(X) = \lambda$. Isso torna a distribuição

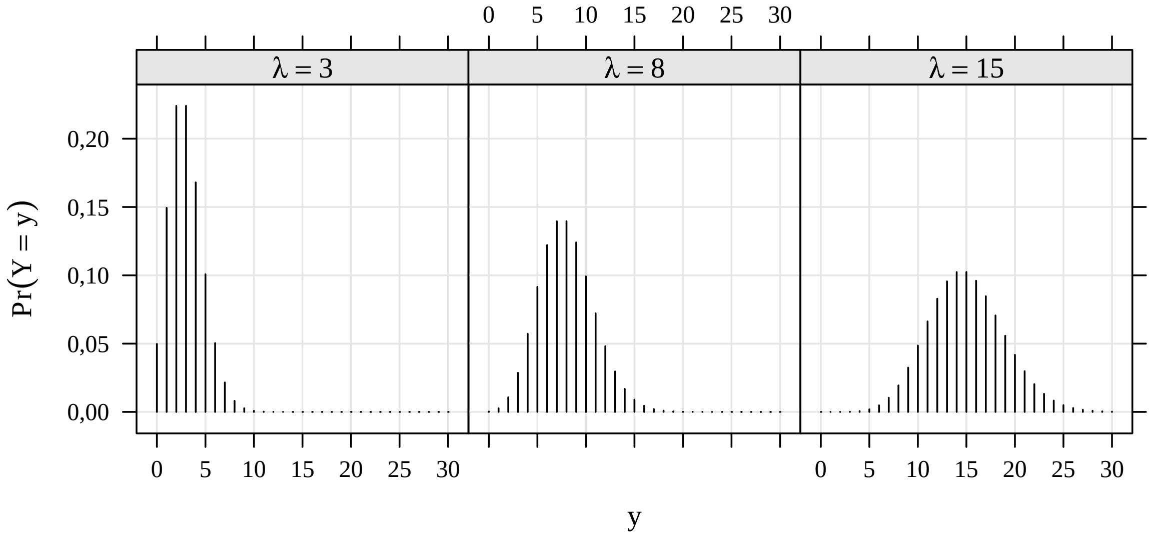


Figura 2 – Probabilidades pela distribuição Poisson para diferentes valores de λ

Poisson bastante reestritiva. 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.

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, $i = 1, 2, \cdots, n$. O modelo de regressão log-linear Poisson, sob a teoria dos MLG's é definido como

$$\begin{aligned} Y_i \mid X_i &\sim Poisson(\mu_i) \\ \log(\mu_i) &= X_i \beta \end{aligned} \tag{2.2}$$

em que $\mu_i > 0$ é a média da variável aleatória $Y_i \mid X_i$ 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 *Iteractive 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 biparamétrica, denoninada quase-verossimilhança. Suponha que temos y_i observações independentes com esperanças μ_i e variâncias $V(\mu_i)$. A função de quase-verossimilhança é é expressa como

$$Q(\mu_i \mid y_i) = \int_y^{\mu_i} \frac{y_i - t}{\phi V(\mu_i)} dt \tag{2.3}$$

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

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

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

ou seja a informação a respeito de μ quando se conhece apenas ϕ 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 $\ell(\mu \mid y)$. Além disso ressalta-se que, de forma geral, não se recupera a distribuição de Y somente com as especificações de ϕ 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 ϕ. A estimativa do vetor β pode ser obtidas pelo algoritmo *IWLS*, usando as funções quase-escore e matriz de quase-informação. Para o parâmetro ϕ um estimador usual é o baseado na estatística χ^2 de Pearson.

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \tag{2.4}$$

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

$$\begin{aligned} Y\,|b &\sim Poisson(b) \\ b &\sim Gama(\mu, \phi) \end{aligned} \tag{2.5}$$

A função massa de probabilidade decorrente da estrutura descrita em 2.6 é deduzida integrando os efeitos aleatórios, considere $f(y \mid b)$ como a função massa de probablidade da distribuição Poisson (vide expressão em 2.1) e $g(b \mid \mu, \phi)$ a função densidade da distribuição Gama [1]

$$\begin{aligned} Pr(Y = y \mid \mu, \phi) &= \int_0^\infty f(y \mid b) g(b \mid \mu, \phi) db \\ &= \frac{\phi^\phi}{y! \mu^\phi \Gamma(\phi)} \int_0^\infty e^{-b(1+\phi/\mu)} b^{y+\phi-1} db \\ &= \frac{\Gamma(\phi + y)}{\Gamma(y+1)\Gamma(\phi)} \left(\frac{\mu}{\mu+\phi}\right)^y \left(\frac{\phi}{\mu+\phi}\right)^\phi \qquad y = 0, 1, 2, \cdots \end{aligned} \tag{2.6}$$

com $\mu > 0$ e $\phi > 0$. 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 ϕ em comparação com a distribuição Poisson equivalente em locação. Note que quanto menor o parâmetro ϕ, maior a dispersão da distribuição. Isso introduz uma propriedade importante desse modelo, para $\phi \to \infty$ a distribuição reduz-se a Poisson.

Os momentos média e variância da distribuição Binomial Negativa são expressos como $E(Y) = \mu$ e $V(Y) = \mu + \mu^2/\phi$. 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 ϕ. Percebemos também quanto maior o parâmetro ϕ mais $E(Y)$ se aproxima de $V(Y)$, e no limite $\phi \to \infty$, $E(Y) = V(Y)$ fazendo com que a distribuição Binomial Negativa se reduza 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.

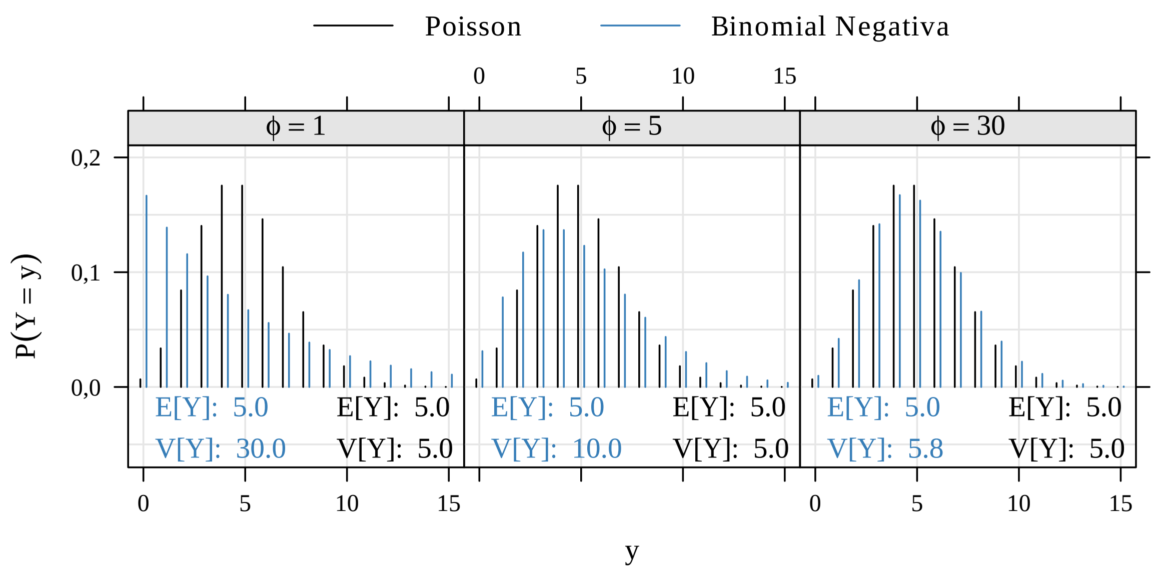


Figura 3 – Probabilidades pela distribuição Binomial Negativa para diferentes valores de ϕ com $\mu = 5$

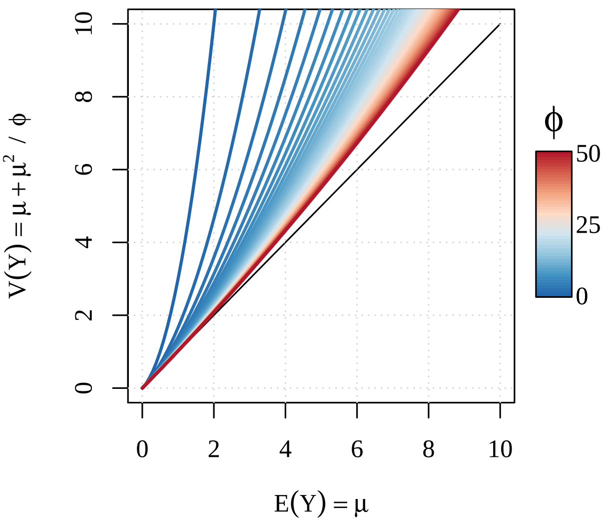


Figura 4 – Relação Média e Variância na distribuição Binomial Negativa

A relação funcional entre média e variância é ilustrada na figura 4 onde apesentamos as médias e variâncias para μ entre 0 e 10 e ϕ entre 0 e 50. O comportamento dessa relação proporciona um mairo 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 ϕ deve ser extremamente grande.

O emprego do modelo Binomial Negativo em problemas se regressão ocorre de maneira similar aos MLG's, com excessão de que a distribuição só pertence a família exponencial de distribuições se o parâmetro ϕ 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 escore e matriz de informação de Fisher chegamos que a matriz de informação é bloco diagonal caracterizando a ortogonalidade dos parâmetros β de locação e ϕ de dispersão. Deste fato decorre que a estimação dos parâmetros pode ser realizada em paralelo, ou seja, estima-se o vetor *beta* pelo método de *IWLS* e posteriormente o parâmetro ϕ pelo método de Newton-Raphson, faz-se os dois procedimentos simultaneamente até a convengência dos parâmetros.

2.3 Modelo COM-Poisson

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

$$Pr(Y = y \mid \lambda, \nu) = \frac{\lambda^y}{(y!)^\nu Z(\lambda, \nu)} \qquad y = 0, 1, 2, \cdots \tag{2.7}$$

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

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu} \tag{2.8}$$

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

$$\frac{Pr(Y = y - 1)}{Pr(Y = y)} = \frac{y^\nu}{\lambda} \tag{2.9}$$

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 consequentemente uma dispersão extra Poisson, ii) contempla subdisersão mesmo em contagens altas, o que na Poisson teriamos 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).

Uma das vantagens do modelo COM-Poisson é que possui, além da Poisson quando $\nu = 1$ outros 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 2.7 temos a distribuição

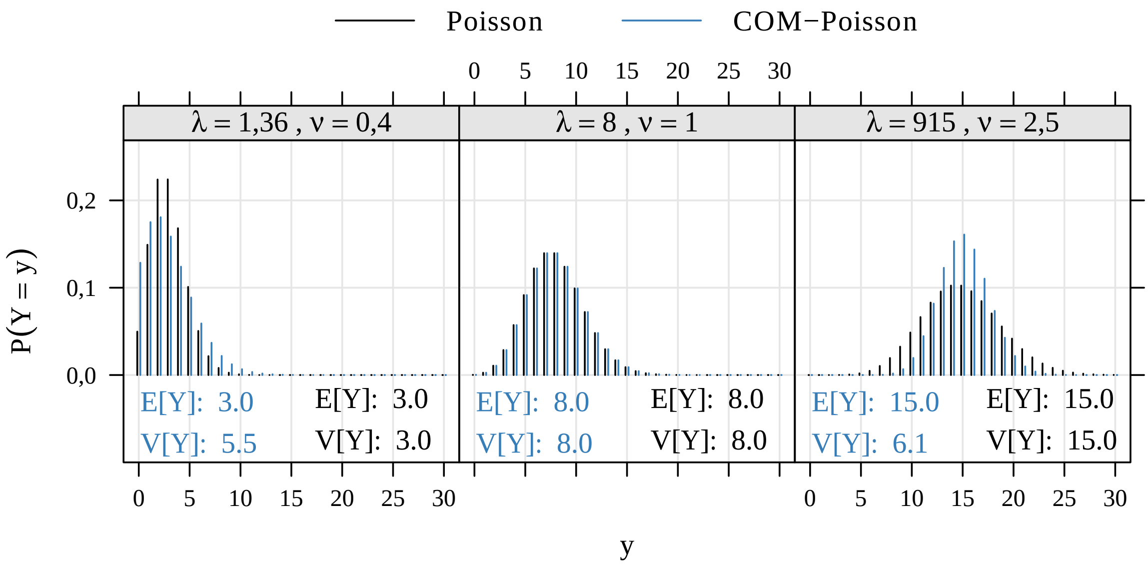


Figura 5 – Probabilidades pela distribuição COM-Poisson para diferentes parâmetros

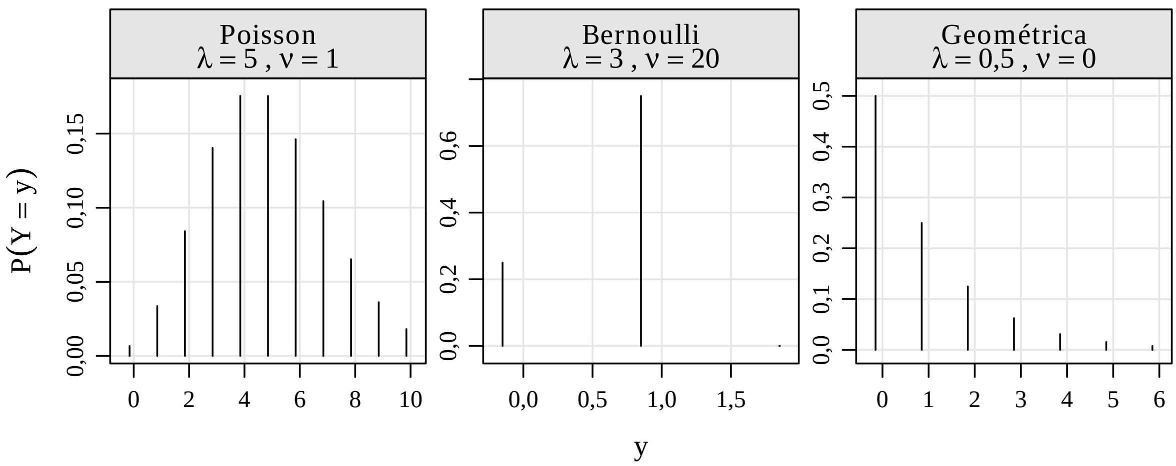


Figura 6 – Exemplos de casos particulares da distribuição COM-Poisson

Poisson resultante. Quando $\nu \to \infty$, $Z(\lambda, \nu) \to 1 + \lambda$ 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 2.7 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 reestrições para redução da distribuição.

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

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

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

$$E(Y) \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu} \qquad\qquad V(Y) \approx \frac{\lambda^{1/\nu}}{\nu} \tag{2.10}$$

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 2.10. 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 $\nu > 1$.

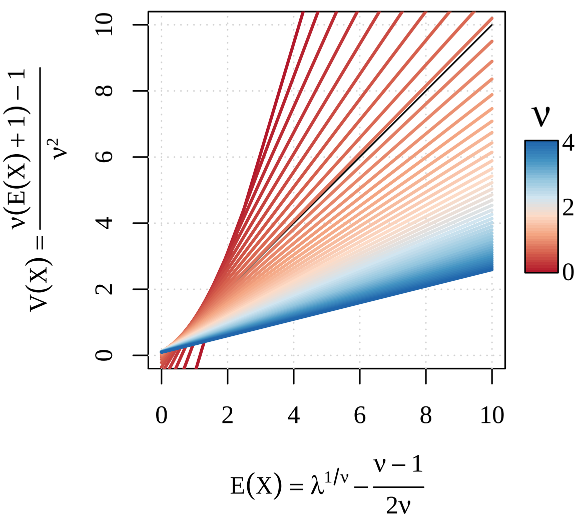


Figura 7 – Relação Média e Variância na distribuição COM-Poisson

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 $Y_1, Y_2, \cdots, Y_n$, dado o vetor de covariáveis $X_i = (x_{i1}, x_{i2}, \cdots, x_{ip})$ seguindo um modelo COM-Poisson de parâmetros $\lambda_i = e^{X_i\beta}$, $i = 1, 2, \cdots, n$ e ν comum a todas as observações. Sob a notação de MLG's, temos em 2.2 o modelo devidamente formulado

$$\begin{aligned} Y_i \mid X_i &\sim \textit{COM-Poisson}(\lambda_i, \nu) \\ \eta(E(Y_i \mid X_i)) &= \log(\lambda_i) = X_i\beta \end{aligned} \tag{2.11}$$

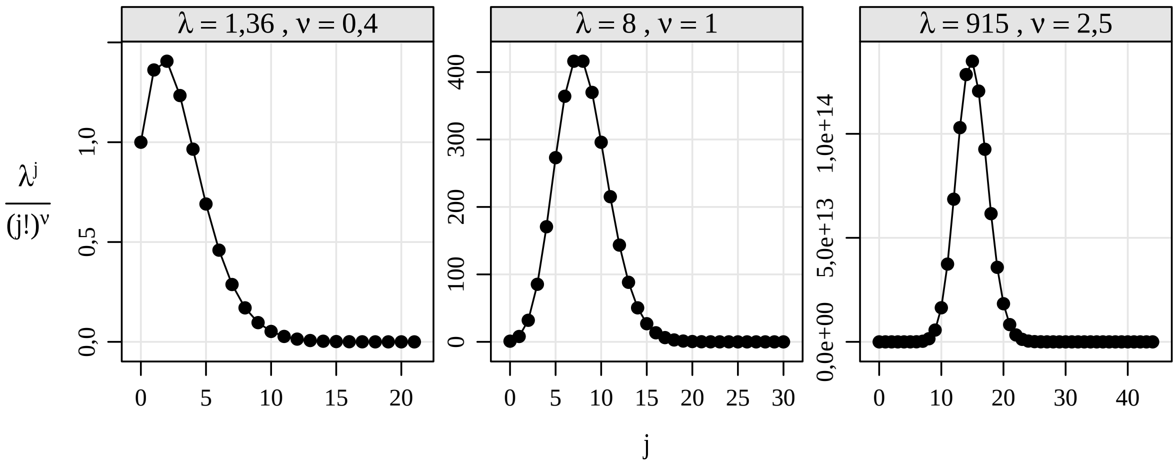


Figura 8 – Convergência da constante de normalização da COM-Poisson para diferentes conjuntos de parâmetros

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 2.11 é

$$\ell(\nu, \beta \mid \underset{\sim}{y}) = \sum_{i}^{n} y_i \log(\lambda_i) - \nu \sum_{i}^{n} \log(y!) - \sum_{i}^{n} \log(Z(\lambda_i, \nu)) \tag{2.12}$$

e então as estimativas de máxima verossimilhança são

$$\hat{\Theta} = (\hat{\nu}, \hat{\beta}) = \underset{(\nu, \beta)}{\text{arg max}}\ \ell(\nu, \beta \mid \underset{\sim}{y})$$

Note que nessa maximização a constante de normalização $Z(\lambda, \nu)$, conforme definida em 2.8 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 convergência [2]. de $Z(\lambda, \nu)$ foram 22, 31, 45 nos primeiro, segundo e terceiro painéis respectivamente.

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

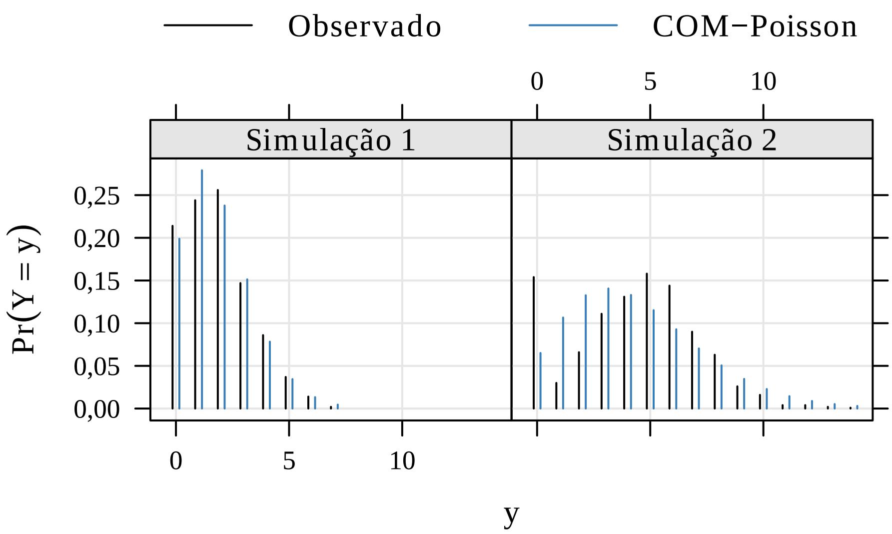


Figura 9 – Ilustração de dados de contagem com excesso de zeros

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.

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

$$Pr(Y = y \mid \pi, \Theta_c) = \begin{cases} \pi & \text{se } \; y = 0, \\ (1-\pi)\dfrac{Pr(Z = z \mid \Theta_c)}{1 - Pr(Z = 0 \mid \Theta_c)} & \text{se } \; y = 1, 2, \ldots \end{cases} \tag{2.13}$$

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

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

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

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = G_i\gamma \qquad e \qquad \begin{array}{c} Z_i \sim D(\mu_i, \phi) \\ g(\mu_i) = X_i\beta \end{array} \tag{2.14}$$

com $i = 1, 2, \cdots, n$, 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 2.14 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

$Y_1, Y_2, \cdots, Y_n$ são condicionalmente indenpendentes, 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

$$\begin{aligned} Y_{ij} \mid b_i, X_{ij} &\sim \mathrm{D}(\mu_{ij}, \phi) \\ g(\mu_{ij}) &= X_{ij}\beta + Z_i b_i \\ b &\sim \mathrm{K}(\Theta_b) \end{aligned} \tag{2.15}$$

para $i = 1, 2, \cdots, m$ (grupos com efeitos aleatórios comuns) e $j = 1, 2, \cdots, n$ (observações) com $\mathrm{D}(\mu_{ij}, \phi)$, uma distribuição considerada para as variáveis resposta condicionalmente independentes, $g(\mu_{ij})$ uma função de ligação conforme definada 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 $\mathrm{K}(\Theta_b)$. 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, $Y \mid X$ e b, a verossimilhança para um modelo de efeito aleatório é dada integrando-se os efeitos aleatórios

$$\mathcal{L}(\beta, \phi, \Theta_b \mid \underline{y}) = \prod_{i=1}^{m} \int_{\mathbb{R}^q} \left(\prod_{j=1}^{n_i} f_D(y_{ij}, \mu, b_i) \right) \cdot f_K(b \mid \Theta_b) db_i \tag{2.16}$$

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 númericos de aproximação, que são discutidos na seção 3.2. E as estimativas de máxima verossimilhança são

$$\hat{\Theta} = (\hat{\beta}, \hat{\Theta}_b) = \underset{(\beta, \Theta_b)}{\arg\max} \; \log(\mathcal{L}(\beta, \phi, \Theta_b \mid \underline{y}))$$

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 2.15 $\mathrm{K}(\Theta_b) = NMV_q(0, \Sigma)$. Para estes casos os principais métodos de aproximação da integral tem desempenhos melhores (**??**).

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 2.16 tem solução analítica e consequentemente a marginal em Y forma fechada.

3 Material e Métodos

3.1 Materias

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer 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.

3.2 Métodos

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer 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.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

4 Resultados e Discussão

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer 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.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer 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,

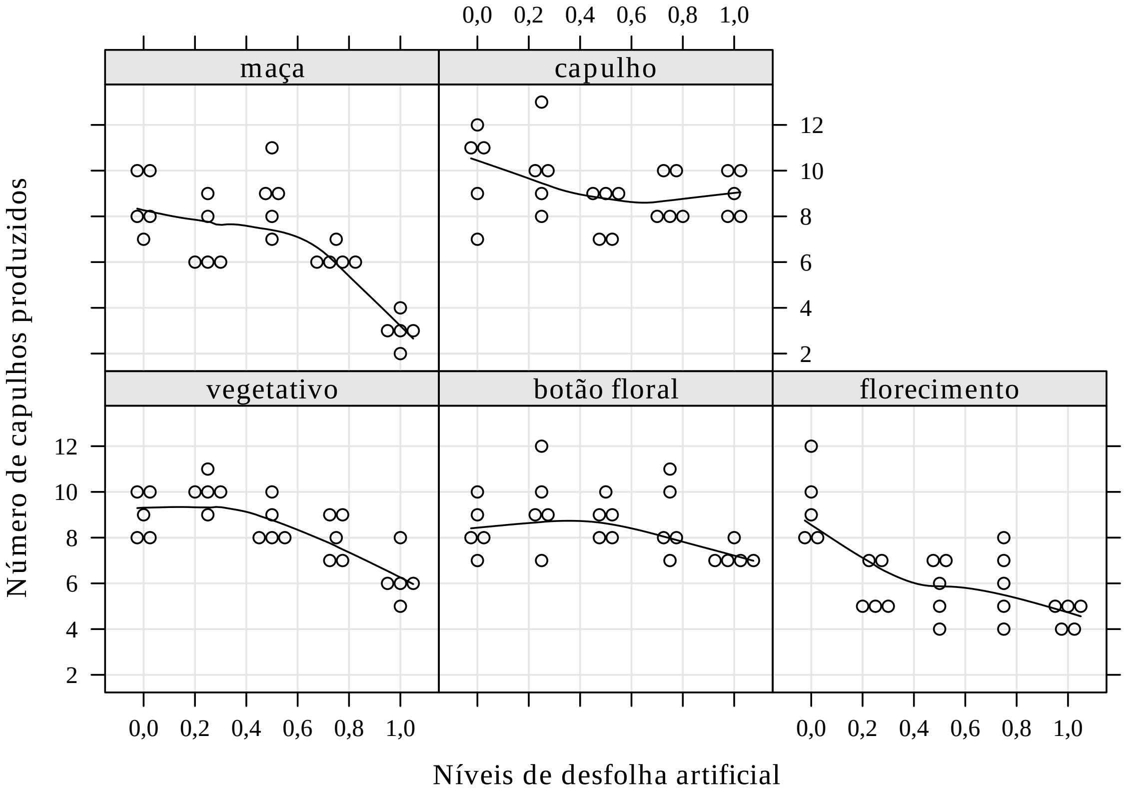


Figura 10 – Experimento sobre capulhos de algodão sob efeito de desfolha

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.

Verificando *outputs* R *inline* 0,333. A referenciação de figuras produzidas em *chunks* R é `fig: + chunkname`, assim como vemos nas figuras 11 e 10. Para as tabelas a referenciação também é simples, conforme tabela 2.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer 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 vestibu-

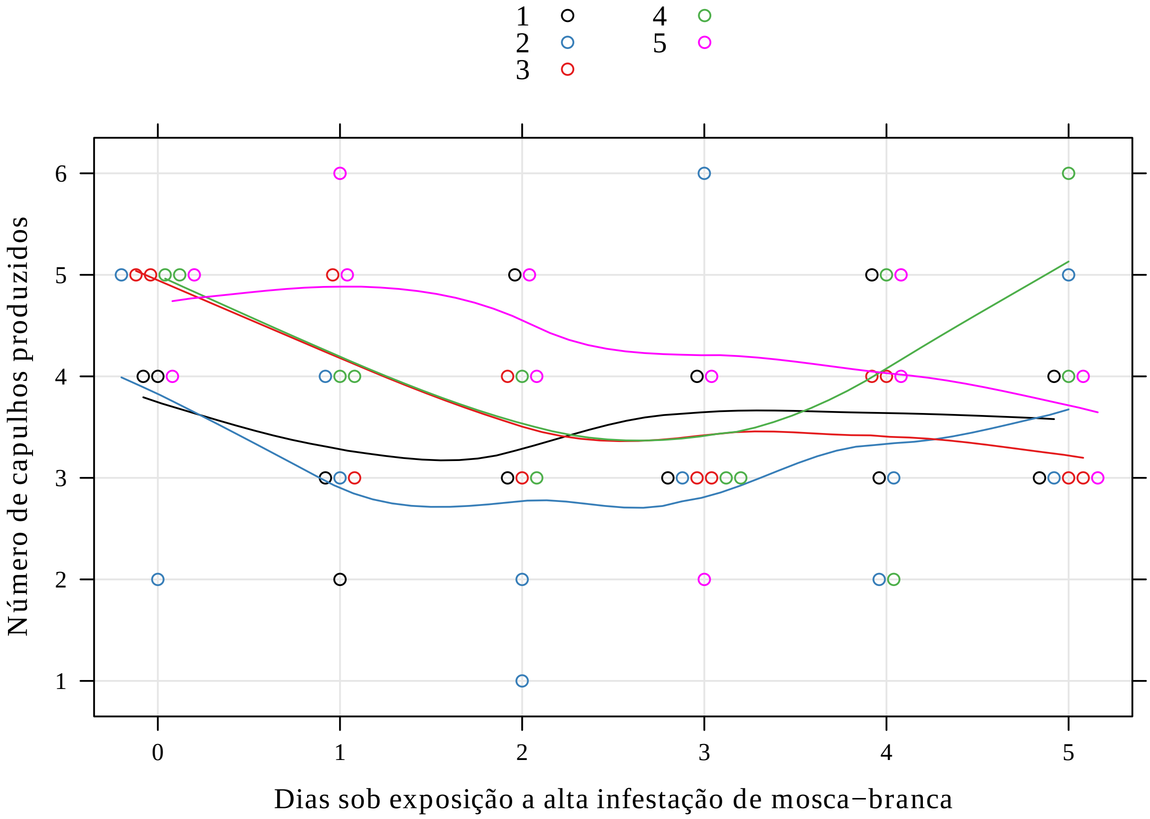


Figura 11 – Experimento sobre capulhos de algodão sob efeito de infestação de mosca-branca

lum 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.

Entendendo citações:

- Zeviani et al. (2014)
- Paula (2013, p. 281)

Tabela 2 – Dados de automóveis

mpg	cyl	disp	hp	drat	wt	qsec	vs
21,00	6,00	160,00	110,00	3,90	2,62	16,46	0,00
21,00	6,00	160,00	110,00	3,90	2,88	17,02	0,00
22,80	4,00	108,00	93,00	3,85	2,32	18,61	1,00
21,40	6,00	258,00	110,00	3,08	3,21	19,44	1,00
18,70	8,00	360,00	175,00	3,15	3,44	17,02	0,00
18,10	6,00	225,00	105,00	2,76	3,46	20,22	1,00
14,30	8,00	360,00	245,00	3,21	3,57	15,84	0,00
24,40	4,00	146,70	62,00	3,69	3,19	20,00	1,00
22,80	4,00	140,80	95,00	3,92	3,15	22,90	1,00
19,20	6,00	167,60	123,00	3,92	3,44	18,30	1,00
17,80	6,00	167,60	123,00	3,92	3,44	18,90	1,00
16,40	8,00	275,80	180,00	3,07	4,07	17,40	0,00
17,30	8,00	275,80	180,00	3,07	3,73	17,60	0,00
15,20	8,00	275,80	180,00	3,07	3,78	18,00	0,00
10,40	8,00	472,00	205,00	2,93	5,25	17,98	0,00
10,40	8,00	460,00	215,00	3,00	5,42	17,82	0,00
14,70	8,00	440,00	230,00	3,23	5,34	17,42	0,00
32,40	4,00	78,70	66,00	4,08	2,20	19,47	1,00
30,40	4,00	75,70	52,00	4,93	1,61	18,52	1,00
33,90	4,00	71,10	65,00	4,22	1,83	19,90	1,00
21,50	4,00	120,10	97,00	3,70	2,46	20,01	1,00
15,50	8,00	318,00	150,00	2,76	3,52	16,87	0,00
15,20	8,00	304,00	150,00	3,15	3,44	17,30	0,00
13,30	8,00	350,00	245,00	3,73	3,84	15,41	0,00
19,20	8,00	400,00	175,00	3,08	3,85	17,05	0,00
27,30	4,00	79,00	66,00	4,08	1,94	18,90	1,00
26,00	4,00	120,30	91,00	4,43	2,14	16,70	0,00
30,40	4,00	95,10	113,00	3,77	1,51	16,90	1,00
15,80	8,00	351,00	264,00	4,22	3,17	14,50	0,00
19,70	6,00	145,00	175,00	3,62	2,77	15,50	0,00
15,00	8,00	301,00	335,00	3,54	3,57	14,60	0,00
21,40	4,00	121,00	109,00	4,11	2,78	18,60	1,00

- (SELLERS; SHMUELI, 2010)
- (Ribeiro Jr et al., 2012, cap. 4)
- (ZEVIANI et al., 2014 apud PARK; LORD, 2009, p. 2–3)
- Lord, Geedipally e Guikema (2010 apud SHMUELI et al., 2005)

5 Considerações Finais

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer 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

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.

HILBE, J. M. *Modeling Count Data*. [S.l.: s.n.], 2014. 300 p. ISSN 1467-9280. ISBN ISBN 978-1-107-02833-3. Citado 2 vezes nas páginas 21 e 31.

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 Underdisperson 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 31.

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 2 vezes nas páginas 21 e 40.

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.

PARK, B.-J.; LORD, D. Application of finite mixture models for vehicle crash data analysis. *Accident; analysis and prevention*, v. 41, n. 4, p. 683–691, 2009. ISSN 1879-2057. Citado na página 40.

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 3 vezes nas páginas 18, 25 e 39.

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 40.

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.2016.01.007http://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 4 vezes nas páginas 21, 27, 29 e 40.

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 4 vezes nas páginas 19, 28, 29 e 40.

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.pdfhttp://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 4 vezes nas páginas 21, 23, 39 e 40.

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, consectetuer 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.