

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

Eduardo Elias Ribeiro Junior
Orientação: Prof. Dr. Walmes Marques Zeviani

Trabalho de Conclusão de Curso - Laboratório B
Departamento de Estatística (DEST)
Universidade Federal do Paraná (UFPR)

28 de junho de 2016

Sumário

1. Introdução
2. Objetivos
3. Materiais e Métodos
4. Resultados e Discussões
5. Considerações finais

1

Introdução

Dados de contagem



São variáveis aleatórias que representam o número de ocorrências de um evento em um domínio discreto ou contínuo.

Se Y é uma variável aleatória de contagem, $y = 0, 1, 2, \dots$

Exemplos:

- ▶ Número de filhos por casal;
- ▶ Número de indivíduos infectados por uma doença;
- ▶ Número de *posts* em uma rede social durante um dia;
- ▶ Número de frutos produzidos;
- ▶ ...

Análise de dados de contagem

- ▶ Modelos de regressão Gaussianos com dados transformados
 - ▶ Dificultam a interpretação dos resultados;
 - ▶ Não contemplam a natureza discreta da variável;
 - ▶ Não contemplam a relação média e variância;
 - ▶ Transformação logarítmica é problemática para valores 0.
- ▶ Modelos de regressão Poisson (NELDER; WEDDERBURN, 1972)
 - ▶ Fiel a natureza dos dados;
 - ▶ Contempla a relação média e variância;
 - ▶ Suposição de equidispersão.

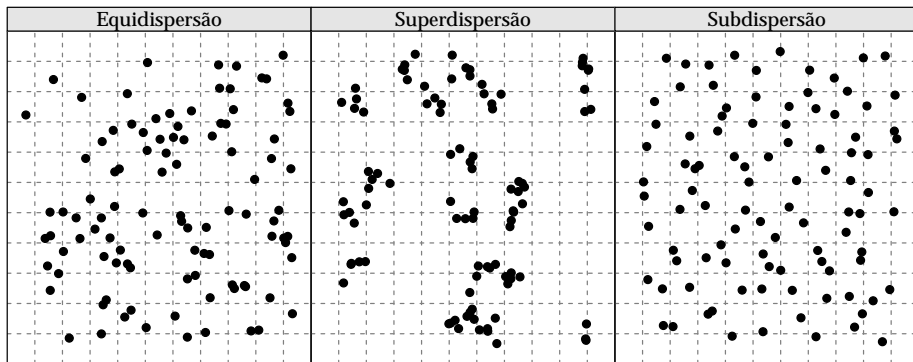


Figura 1: Ilustração de processos pontuais que levam a contagens com diferentes níveis de dispersão.

Distribuições de probabilidades para dados de contagem

Com base em WINKELMANN (2008) e KOKONENDJI (2014)

Tabela 1: Distribuições de probabilidades para dados de contagem

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>	✓	✓	✓

Distribuições de probabilidades para dados de contagem

Com base em WINKELMANN (2008) e KOKONENDJI (2014)

Tabela 1: Distribuições de probabilidades para dados de contagem

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>	✓	✓	✓

Distribuição COM-Poisson

Proposta por CONWAY; MAXWELL (1962).

Função massa de probabilidade

$$\Pr(Y = y \mid \lambda, \nu) = \frac{\lambda^y}{(y!)^\nu Z(\lambda, \nu)}, \quad Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu} \quad (1)$$

- ▶ Não tem expressão fechada para média e variância;
- ▶ Apresenta distribuições bastante conhecidas como casos particulares:
 - ▶ Poisson - $\nu = 1$;
 - ▶ Bernoulli - $\nu \rightarrow \infty$;
 - ▶ Geométrica - $\nu = 0$ e $\lambda < 1$.

Distribuição COM-Poisson

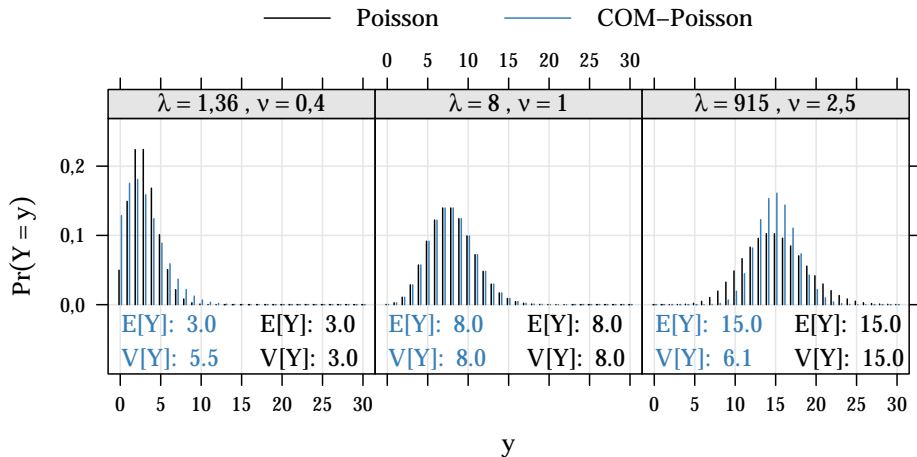


Figura 2: Probabilidades pela distribuição COM-Poisson.

Relações média-variância

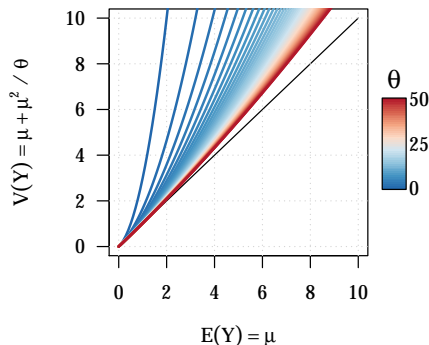
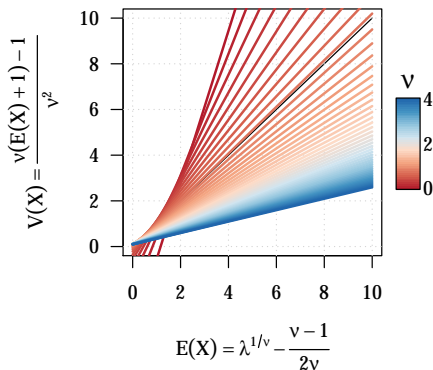


Figura 3: Relações Média e Variância COM-Poisson e Binomial Negativa.

2

Objetivos

Objetivos gerais

Colaborar com a literatura estatística brasileira, no que diz respeito a dados de contagem:

- ▶ Apresentando e explorando o modelo de regressão COM-Poisson;
- ▶ Estendendo o modelo para modelagem de excesso de zeros e inclusão de efeitos aleatórios;
- ▶ Discutindo o desempenho do modelo via análise de dados reais;
- ▶ Disponibilizando os recursos computacionais para ajuste dos modelos, em formato de pacote R.

3

Materiais e Métodos

3.1

Materiais e Métodos
Materiais

Conjuntos de dados

Seis conjuntos de dados analisados:

- ▶ Capulhos de algodão sob desfolha artificial;
- ▶ Produtividade de algodão sob infestação de Mosca-branca;
- ▶ Produtividade de soja sob umidade e adubação potássica;
- ▶ Ocorrência de ninfas de Mosca-branca em lavoura de soja;
- ▶ Peixes capturados por visitantes de um parque Estadual;
- ▶ Número de nematoides em raízes de feijoeiro.

Conjuntos de dados

Seis conjuntos de dados analisados:

- ▶ Capulhos de algodão sob desfolha artificial;
- ▶ Produtividade de algodão sob infestação de Mosca-branca;
- ▶ Produtividade de soja sob umidade e adubação potássica;
- ▶ Ocorrência de ninfas de Mosca-branca em lavoura de soja;
- ▶ Peixes capturados por visitantes de um parque Estadual;
- ▶ Número de nematoides em raízes de feijoeiro.

Recursos Computacionais

Software R versão 3.3.0. Principais pacotes:

- ▶ MASS - 7.3.45: ajuste dos modelos binomial negativo;
- ▶ pscl - 1.4.9: modelagem de excesso de zeros;
- ▶ lme4 - 1.1.12: ajuste dos modelos Poisson com efeito aleatório Normal;
- ▶ bbmle - 1.0.18: ajuste de modelos via máxima verossimilhança.

3.2

Materiais e Métodos
Métodos

Estimação via máxima verossimilhança

- 1 Escreva a função de verossimilhança - $\mathcal{L}(\Theta \mid \underline{y})$
- 2 Tome seu logaritmo - $\ell(\Theta \mid \underline{y})$
- 3 As estimativas dos parâmetros são

$$\hat{\Theta} = \arg \max_{\Theta} \ell(\Theta \mid \underline{y})$$

- ▶ Algoritmo IWLS (*Interactive Weighed Leasts Squares*) para os modelos Poisson, Binomial Negativo e Quasi-Poisson.
- ▶ Método *BFGS* para os modelos COM-Poisson.

Verossimilhança do modelo COM-Poisson

- ▶ Reparametrizando $\phi = \log(\nu)$
 - ▶ $\phi < 0 \Rightarrow$ Superdispersão
 - ▶ $\phi = 0 \Rightarrow$ Equidispersão
 - ▶ $\phi > 0 \Rightarrow$ Subdispersão

Log-verossimilhança

$$\ell(\phi, \beta \mid \underline{y}) = \sum_{i=1}^n y_i \log(\lambda_i) - e^{\phi} \sum_{i=1}^n \log(y_i!) - \sum_{i=1}^n \log(Z(\lambda_i, \phi)) \quad (2)$$

em que $\lambda_i = e^{X_i \beta}$, com X_i o vetor $(x_{i1}, x_{i2}, \dots, x_{ip})$ de covariáveis da i -ésima observação, e $(\beta, \phi) \in \mathbb{R}^{p+1}$.

Verossimilhança do modelo Hurdle COM-Poisson

- ▶ $\underline{\pi} = \frac{\exp(G\gamma)}{1+\exp(G\gamma)}$ a probabilidade de contagem nula.
- ▶ $\underline{\lambda} = \exp(X\beta)$ o parâmetro de locação da distribuição COM-Poisson truncada.

Verossimilhança

$$\mathcal{L}(\phi, \beta, \gamma \mid \underline{y}) = \mathbb{1}[\underline{\pi}] \cdot (1 - \mathbb{1}) \left[(1 - \underline{\pi}) \left(\frac{\underline{\lambda}^y}{(y!)^{e^\phi} Z(\underline{\lambda}, \phi)} \right) \left(1 - \frac{1}{Z(\underline{\lambda}, \phi)} \right) \right] \quad (3)$$

em que $\mathbb{1}$ é uma função indicadora para $y = 0$

Verossimilhança do modelo misto COM-Poisson

$$Y_{ij} \mid b_i, X_{ij} \sim \text{COM-Poisson}(\mu_{ij}, \phi)$$

$$g(\mu_{ij}) = X_{ij}\beta + Z_i b_i$$

$$b \sim \text{Normal}(0, \Sigma)$$

Verossimilhança

$$\mathcal{L}(\phi, \Sigma, \beta \mid \underline{y}) = \prod_{i=1}^m \int_{\mathbb{R}^q} \left(\prod_{j=1}^{n_i} \frac{\underline{\lambda}^y}{(y!)^{e^\phi} Z(\underline{\lambda}, \phi)} \right) \cdot (2\pi)^{q/2} |\Sigma| \exp \left(-\frac{1}{2} b^t \Sigma^{-1} b \right) db_i \quad (4)$$

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.

4

Resultados e Discussões

4.1

Resultados e Discussões

Pacote R

cmpreg: Ajuste de Modelos de Regressões COM-Poisson

Implementação em R de um *framework* para ajuste dos modelos de regressão COM-Poisson, pacote cmpreg.

```
## Pode ser instalado do GitHub
devtools::install_git("https://github.com/JrEduardo/cmpreg.git")
library(cmpreg)

## Regressão (efeitos fixos)
cmp(y ~ predictor, data = data)

## Regressão com componente de barreira
hurdlecmp(y ~ count_pred | zero_pred, data = data)

## Regressão (efeitos aleatórios)
mixedcmp(y ~ count_pred + (1 | ind.ranef), data = data)
```

4.2

Resultados e Discussões
Produtividade de algodão

Experimento

Conduzido na UFGD em casa de vegetação (MARTELLI et al., 2008).

- ▶ Objetivo: avaliar o impacto da praga Mosca-branca na produção de algodão;
- ▶ Delineamento: inteiramente casualizado com cinco repetições
- ▶ Unidade amostral: vaso com duas plantas;
- ▶ Covariável experimental:
 - ▶ Tempo de exposição das plantas à praga, em dias (dexp);
- ▶ Variáveis resposta:
 - ▶ Número de capulhos produzidos;
 - ▶ Número de estruturas reprodutivas;
 - ▶ Número de nós.

Modelagem

Preditores considerados:

- ▶ Preditor 1: $g(\mu_i) = \beta_0$
- ▶ Preditor 2: $g(\mu_i) = \beta_0 + \beta_1 \text{dexp}_i$
- ▶ Preditor 3: $g(\mu_i) = \beta_0 + \beta_1 \text{dexp}_i + \beta_2 \text{dexp}_i^2$

Modelos concorrentes:

- ▶ Poisson(μ_i)
- ▶ COM-Poisson(λ_i, ϕ)
- ▶ Quasi-Poisson(μ_i, σ^2)

Medidas de ajuste

Tabela 2: Medidas de ajuste para avaliação e comparação

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	0,4832	-91,31	188,62	0,2254	20,31	0,2296
3	-104,44	214,88	0,2782	-89,47	186,95	0,0552	19,13	0,0616
Número de estruturas reprodutivas								
1	-104,74	211,49		-86,41	176,82		16,23	
2	-104,27	212,54	0,3320	-84,59	175,18	0,0566	15,29	0,0622
3	-104,06	214,12	0,5157	-83,73	175,47	0,1898	14,87	0,2071
Número de nós da planta								
1	-143,79	289,59		-120,58	245,16		12,69	
2	-143,48	290,95	0,4253	-119,03	244,06	0,0787	12,05	0,0851
3	-142,95	291,89	0,3037	-116,27	240,54	0,0188	11,00	0,0223

Medidas de ajuste

Tabela 2: Medidas de ajuste para avaliação e comparação

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	0,4832	-91,31	188,62	0,2254	20,31	0,2296
3	-104,44	214,88	0,2782	-89,47	186,95	0,0552	19,13	0,0616
Número de estruturas reprodutivas								
1	-104,74	211,49		-86,41	176,82		16,23	
2	-104,27	212,54	0,3320	-84,59	175,18	0,0566	15,29	0,0622
3	-104,06	214,12	0,5157	-83,73	175,47	0,1898	14,87	0,2071
Número de nós da planta								
1	-143,79	289,59		-120,58	245,16		12,69	
2	-143,48	290,95	0,4253	-119,03	244,06	0,0787	12,05	0,0851
3	-142,95	291,89	0,3037	-116,27	240,54	0,0188	11,00	0,0223

Avaliação da dispersão

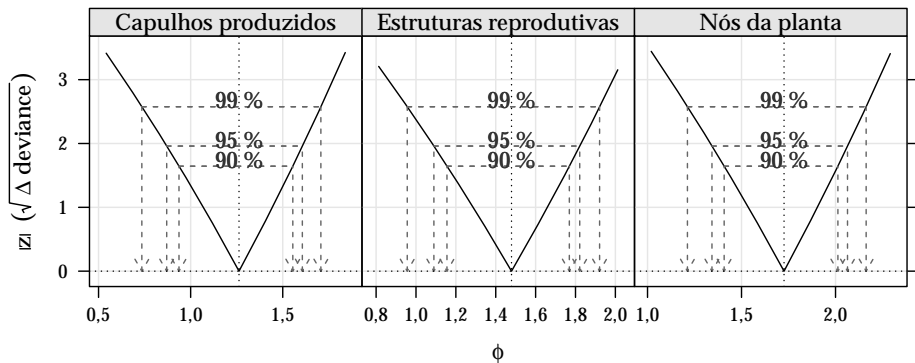


Figura 4: Perfis de log-verossimilhança para o parâmetro de precisão da COM-Poisson.

Avaliação da matriz de covariância

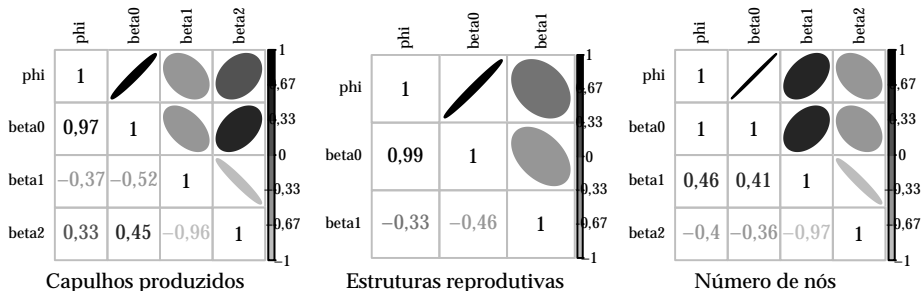


Figura 5: Imagem da matriz de correlação entre os parâmetros do modelo COM-Poisson.

Valores preditos

----- Poisson — COM-Poisson - - - - Quasi-Poisson

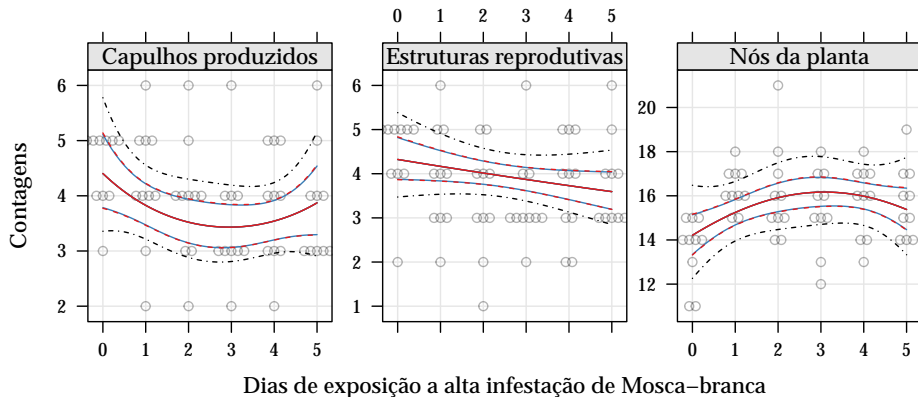


Figura 6: Curva dos valores preditos com intervalo de confiança de (95%) como função dos dias de exposição a alta infestação de Mosca-branca.

4.3

Resultados e Discussões
**Ocorrência de ninfas de
Mosca-branca**

Experimento

Conduzido na UFGD em casa de vegetação (SUEKANE, 2011).

- ▶ Objetivo: avaliar a ocorrência de mosca-branca nas diferentes cultivares de soja;
- ▶ Delineamento: blocos casualizados, quatro blocos;
- ▶ Unidade experimental: dois vasos com duas plantas;
- ▶ Covariáveis experimentais:
 - ▶ Indicadora de bloco, I, II, III e IV, (bloco);
 - ▶ Dias decorridos após a primeira avaliação, 0, 8, 13, 22, 31 e 38 dias, (dias);
 - ▶ Indicadora de cultivar de soja, BRS 239, BRS 243 RR, BRS 245 RR, BRS246 RR, (cult);
- ▶ Variável resposta:
 - ▶ Número de ninfas de Mosca-branca nos folíolos dos terços superior, médio e inferior.

Modelagem

Preditores considerados:

- ▶ Preditor 1: $g(\mu_{ijk}) = \beta_0 + \tau_i + \gamma_j + \delta_k$
- ▶ Preditor 2: $g(\mu_{ijk}) = \beta_0 + \tau_i + \gamma_j + \delta_k + \alpha_{jk}$

τ_i é o efeito do i-ésimo bloco, $i = 1, 2, 3, 4$

γ_j o efeito da j-ésima cultivar, $j = 1, 2, 3, 4$

δ_k o efeito do k-ésimo nível de dias, $k = 1, 2, \dots, 6$ e

α_{jk} o efeito da interação entre a j-ésima cultivar e o k-ésimo nível de dias

Modelos concorrentes:

- ▶ Poisson(μ_{ijk})
- ▶ COM-Poisson(λ_{ijk}, ϕ)
- ▶ Binomial Negativo(μ_{ijk}, θ)
- ▶ Quasi-Poisson(μ_{ijk}, σ^2)

Medidas de ajuste

Tabela 3: Medidas de ajuste para avaliação e comparação

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	0,0000	
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	0,9680	-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	0,7376	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	0,9932	19,03

Medidas de ajuste

Tabela 3: Medidas de ajuste para avaliação e comparação

Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	P(> χ^2)	
Preditor 1	12	-922,98	1869,96				
Preditor 2	27	-879,23	1812,46	87,50	15	0,0000	
COM-Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	P(> χ^2)	$\hat{\phi}$
Preditor 1	13	-410,44	846,89				-3,08
Preditor 2	28	-407,15	870,30	6,59	15	0,9680	-2,95
Binomial Neg.	np	ℓ	AIC	2(diff ℓ)	diff np	P(> χ^2)	$\hat{\theta}$
Preditor 1	13	-406,16	838,31				3,44
Preditor 2	28	-400,55	857,10	11,21	15	0,7376	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	0,9932	19,03

Avaliando a dispersão e convergência de Z

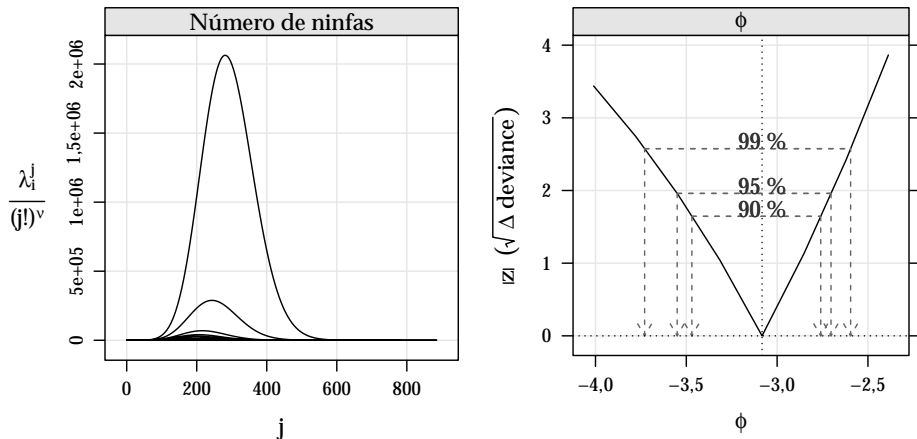


Figura 7: Convergência das constantes de normalização e perfil de log-verossimilhança para o parâmetro de precisão da COM-Poisson.

Valores preditos

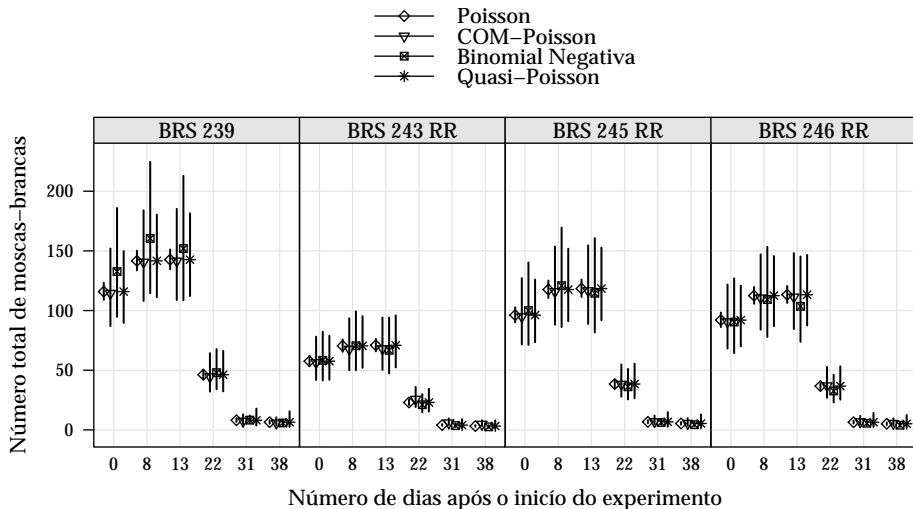


Figura 8: Valores preditos com intervalos de confiança (95%).

4.4

Resultados e Discussões
Peixes capturados

Estudo

Observacional conduzido por biólogos em um Parque Estadual (UCLA, 2015).

- ▶ Delineamento: amostragem aleatória.
- ▶ Objetivo: modelar o número de peixes capturados pela atividade de pesca esportiva.
- ▶ Unidade experimental: grupos de pescadores visitantes do parque.
- ▶ Covariáveis mensuradas:
 - ▶ Número de pessoas, (n_p),
 - ▶ Número de crianças. (n_c),
 - ▶ Indicador de campista no grupo, (ca).
- ▶ Variável resposta:
 - ▶ Número de peixes capturados pelo grupo.

Modelagem

Preditores considerados:

- ▶ Preditor 1:
$$g(\mu_i) = \beta_0 + \beta_1 ca_i + \beta_2 np_i$$
$$\text{logit}(\pi_i) = \gamma_0 + \gamma_1 ca_i + \gamma_2 np_i + \gamma_3 nc_i$$
- ▶ Preditor 2:
$$g(\mu_i) = \beta_0 + \beta_1 ca_i + \beta_2 np_i + \beta_3 nc_i + \beta_4 (np_i \cdot nc_i)$$
$$\text{logit}(\pi_i) = \gamma_0 + \gamma_1 ca_i + \gamma_2 np_i + \gamma_3 nc_i + \gamma_4 (np_i \cdot nc_i)$$

Modelos concorrentes:

- ▶ Hurdle Poisson(π_i, μ_i)
- ▶ Hurdle COM-Poisson(π_i, λ_i, ϕ)
- ▶ Hurdle Binomial Negativo(π_i, μ_i, θ)

Medidas de ajuste

Tabela 4: Medidas de ajuste para avaliação e comparação

Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$	
Preditor 1	7	-857,48	1728,96				
Preditor 2	10	-744,58	1509,17	225,79	3	1,1E-48	
Binomial Neg.	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$	$\hat{\theta}$
Preditor 1	8	-399,79	815,58				0,20
Preditor 2	11	-393,72	809,44	12,14	3	0,0069	0,37
COM-Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$	$\hat{\phi}$
Preditor 1	8	-409,85	835,71				-8,77
Preditor 2	11	-402,30	826,59	15,12	3	0,0017	-3,77

Medidas de ajuste

Tabela 4: Medidas de ajuste para avaliação e comparação

Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$	
Preditor 1	7	-857,48	1728,96				
Preditor 2	10	-744,58	1509,17	225,79	3	1,1E-48	
Binomial Neg.	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$	$\hat{\theta}$
Preditor 1	8	-399,79	815,58				0,20
Preditor 2	11	-393,72	809,44	12,14	3	0,0069	0,37
COM-Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$	$\hat{\phi}$
Preditor 1	8	-409,85	835,71				-8,77
Preditor 2	11	-402,30	826,59	15,12	3	0,0017	-3,77

Valores preditos

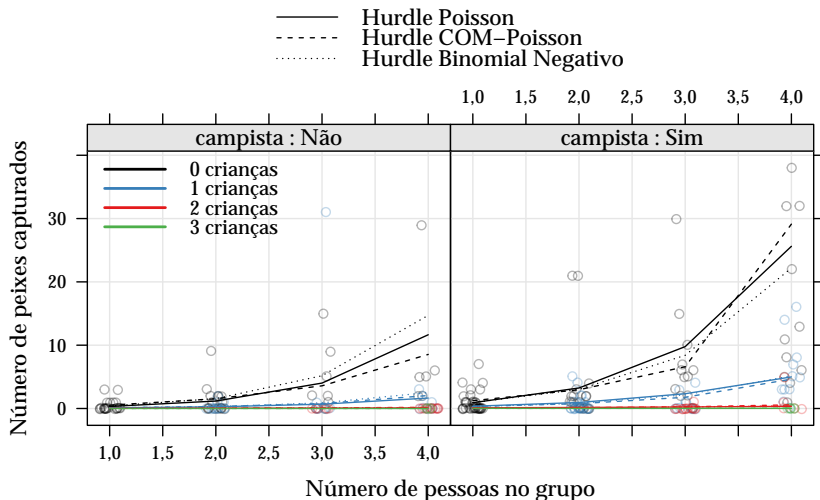


Figura 9: Valores preditos do número de peixes capturados.

4.5

Resultados e Discussões
Número de nematoides

Experimento

Conduzido no IAPAR em casa de vegetação.

- ▶ Objetivo: avaliar a resistência de linhagens de feijoeiro à nematoides;
- ▶ Delineamento: inteiramente casualizado com cinco repetições;
- ▶ Unidade amostral: alíquota de 1ml da solução de raízes lavadas, trituradas, peneiradas, diluídas em água. Provida por um vaso com duas plantas;
- ▶ Covariáveis:
 - ▶ Indicador de linhagem de feijoeiro, A, B, C, ..., S, (cult);
 - ▶ Concentração de raiz na solução, (sol);
- ▶ Variáveis resposta:
 - ▶ Número de nematoides.

Modelagem

Preditores considerados:

- ▶ Preditor 1: $g(\mu_{ij}) = \beta_0 + b_i$
- ▶ Preditor 2: $g(\mu_{ij}) = \beta_0 + \beta_1 \log(\text{sol})_{ij} + b_i$

$$b_i \sim \text{Normal}(0, \sigma^2)$$

i : varia entre as linhagens, $i = 1, 2, \dots, 19$; e

j : varia entre as observações dentro das linhagens, $j = 1, 2, \dots, n_i$.

Modelos concorrentes:

- ▶ Poisson(μ_{ij})
- ▶ COM-Poisson(λ_{ij}, ϕ)

Medidas de ajuste

Tabela 5: Medidas de ajuste para avaliação e comparação

Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$		
Preditor 1	2	-237,20	478,40					
Preditor 2	3	-234,00	474,00	6,40	1	0,0114		
COM-Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$	$\hat{\phi}$	$P(> \chi^2)$
Preditor 1	3	-236,85	479,71				0,15	0,4060
Preditor 2	4	-233,16	474,31	7,40	1	0,0065	0,24	0,1935

Medidas de ajuste

Tabela 5: Medidas de ajuste para avaliação e comparação

Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$		
Preditor 1	2	-237,20	478,40					
Preditor 2	3	-234,00	474,00	6,40	1	0,0114		
COM-Poisson	np	ℓ	AIC	2(diff ℓ)	diff np	$P(> \chi^2)$	$\hat{\phi}$	$P(> \chi^2)$
Preditor 1	3	-236,85	479,71				0,15	0,4060
Preditor 2	4	-233,16	474,31	7,40	1	0,0065	0,24	0,1935

Avaliação dos perfis de verossimilhança

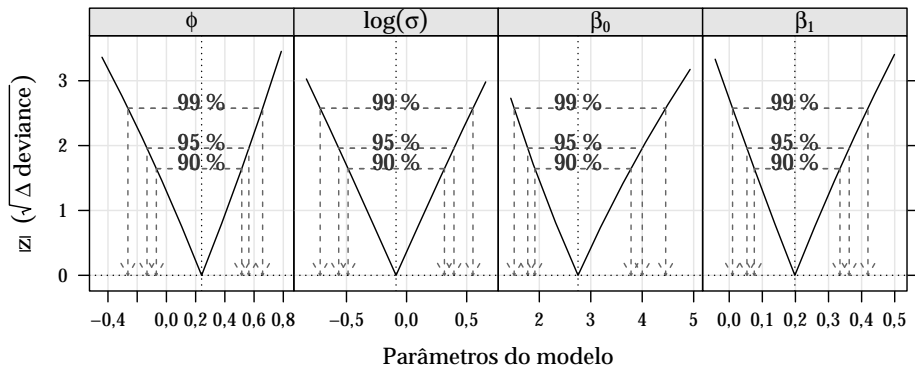


Figura 10: Perfis de verossimilhança dos parâmetros estimados no modelo COM-Poisson Misto.

Imagem da matriz de covariância

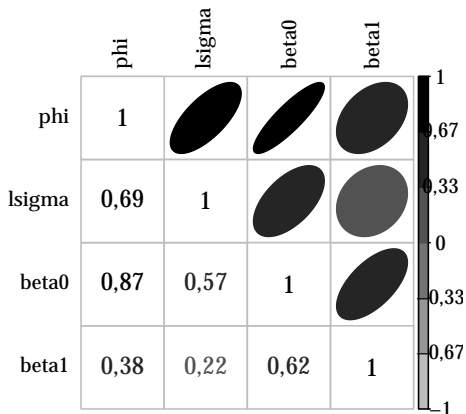


Figura 11: Imagem da matriz de covariância entre os parâmetros do modelo COM-Poisson.

Valores preditos

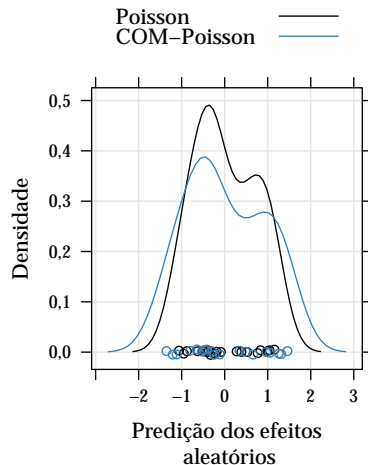
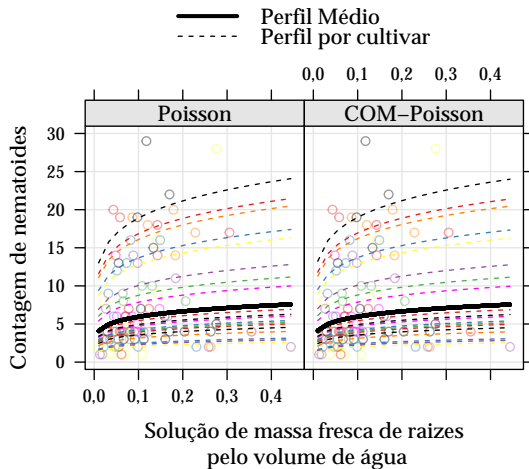


Figura 12: Valores preditos nos modelos de efeitos mistos.

4.6

Resultados e Discussões

Discussões

- ▶ Similaridade entre inferências via modelo Quasi-Poisson e COM-Poisson;
- ▶ Desempenho do modelo Binomial Negativo;
- ▶ Interpretação dos parâmetros nos modelos baseados na COM-Poisson;
- ▶ Problemas numéricos para determinação da matriz hessiana no modelo Hurdle COM-Poisson;
- ▶ Procedimentos computacionalmente intensivos na avaliação da verossimilhança no caso COM-Poisson de efeitos aleatórios;
- ▶ Não ortogonalidade observada (empírica) entre os parâmetros de locação e de precisão no modelo COM-Poisson; e
- ▶ Comportamento simétrico dos perfis de log-verossimilhança para o parâmetro ϕ da COM-Poisson.

5

Considerações finais

Conclusões

Aplicação do modelo COM-Poisson:

- ▶ Resultados similares aos providos pela abordagem semi-paramétrica via quasi-verossimilhança;
- ▶ A não ortogonalidade entre os parâmetros de locação e precisão no modelo COM-Poisson se mostra como característica da distribuição;
- ▶ A simetria nos perfis de verossimilhança do parâmetro de precisão também; e
- ▶ A avaliação da constante de normalização é uma dificuldade computacional do modelo.

Conclusões

Análise de dados de contagem:

- ▶ Modelo Poisson inadequado na maioria das aplicações, mostrando que a suposição de equidispersão é de fato restritiva;
- ▶ Modelos alternativos ao Poisson devem ser empregados na análise de dados de contagem; e
- ▶ Sugere-se o modelo COM-Poisson como alternativa totalmente paramétrica e bastante flexível.

Trabalhos futuros

- ▶ Estudar reparametrizações do modelo COM-Poisson;
- ▶ Avaliar aproximações da constante de normalização;
- ▶ Realizar estudos de simulação para avaliar a robustez do modelo;
- ▶ Implementar o modelo COM-Poisson inflacionado de zeros; e
- ▶ Implementar o modelo COM-Poisson com efeitos aleatórios dependentes.

Publicização



<https://github.com/JrEduardo/cmpreg>

<https://github.com/JrEduardo/tccDocument>



Referências

- CONWAY, R. W.; MAXWELL, W. L. A queuing model with state dependent service rates. **Journal of Industrial Engineering**, v. 12, p. 132—136, 1962.
- KOKONENDJI, C. C. Over- and Underdispersion Models. In: **Methods and applications of statistics in clinical trials: Planning, analysis, and inferential methods**. Tradução. [s.l.: s.n.]. p. 506–526.
- MARTELLI, T. et al. **Influência do ataque de mosca-branca Bemisia tabaci Biotipo B, nos índices de produtividade do algodoeiro** Uberlândia- MGXXII Congresso Brasileiro de Entomologia, 2008.
- 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.
- SUEKANE, R. **DISTRIBUIÇÃO ESPACIAL E DANO DE MOSCA-BRANCA Bemisia tabaci (GENNADIUS, 1889) BIÓTIPO B NA SOJA**. PhD thesis—[s.l.] Universidade Federal da Grande Dourados, 2011.
- UCLA, S. C. G. **Data Analysis Examples**, 2015. Disponível em: <<http://www.ats.ucla.edu/stat/dae/>>
- WINKELMANN, R. **Econometric Analysis of Count Data**. Tradução. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. p. 342