

Extensões e Aplicações do Modelo 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)

17 de agosto de 2016

Sumário

1. Introdução
2. Materiais e Métodos
3. Resultados
4. Extensões
5. Conclusões

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 em Medicina:

- ▶ Número de ocorrências de uma doença por município em um ano;
- ▶ Número de admissões em um hospital por mês;
- ▶ Número de linfócitos em 1mm^3 de sangue.
- ▶ ...

Abordagens comuns para análise

- ▶ 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 ($E(X) = V(X)$).
 - ▶ Produz erros padrões inconsistentes (WINKELMANN; ZIMMERMANN, 1994).

Dispersão em dados de contagem

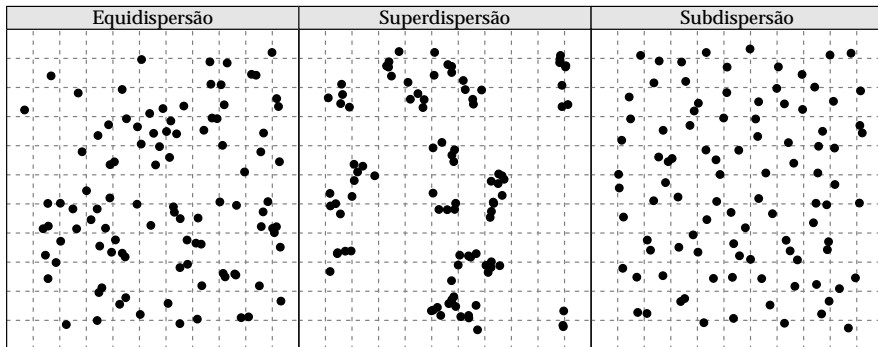


Figura : Ilustração de processos pontuais que levam a contagens com diferentes níveis de dispersão.

1.1

Introdução

Distribuições de probabilidades

Tabela : Distribuições de probabilidades para dados de contagem

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	✓	✓	✓

Tabela : Distribuições de probabilidades para dados de contagem

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	✓	✓	✓

Modelo Poisson

Função massa de probabilidade

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

Propriedades

- ▶ $\frac{P(Y=y-1)}{P(Y=y)} = \frac{y}{\lambda}$
- ▶ $E(Y) = \lambda$
- ▶ $V(Y) = \lambda$

Modelo Poisson

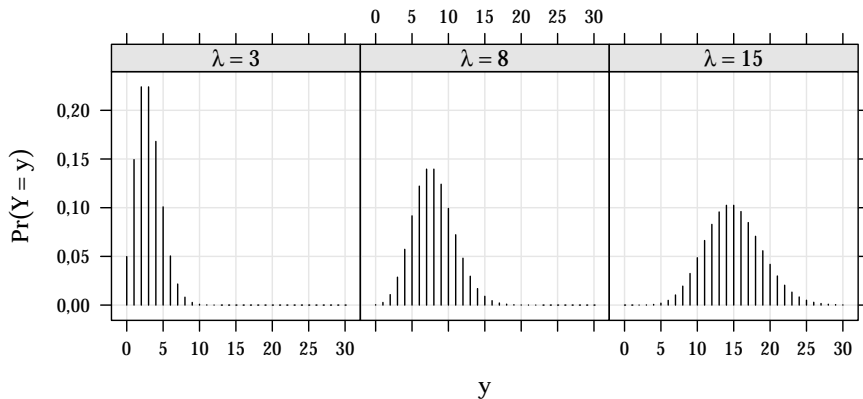


Figura : Probabilidades pela distribuição Poisson para diferentes parâmetros.

Modelo Binomial Negativo

Função massa de probabilidade

$$\Pr(Y = y \mid \mu, \theta) = \frac{\Gamma(\theta + y)}{\Gamma(y + 1)\Gamma(\theta)} \left(\frac{\mu}{\mu + \theta} \right)^y \left(\frac{\theta}{\mu + \theta} \right)^\theta, \quad y = 0, 1, 2, \dots \quad (2)$$

Propriedades

- ▶ $E(Y) = \mu$
- ▶ $V(Y) = \mu + \mu^2/\theta$

Casos particulares

- ▶ Aproximadamente Poisson, quando $\theta \rightarrow \infty$
- ▶ Distribuição Geométrica, quando $\theta = 1$

Modelo Binomial Negativo

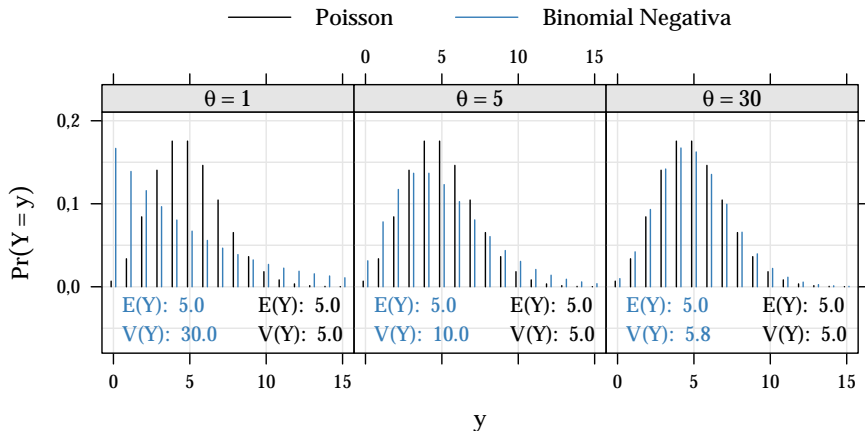


Figura : Probabilidades pela distribuição Binomial Negativa para diferentes níveis de dispersão, fixando a média em 5.

Modelo COM-Poisson

Função massa de probabilidade

$$\Pr(Y = y \mid \lambda, \nu) = \frac{\lambda^y}{(y!)^\nu Z(\lambda, \nu)} \quad y \in \mathbb{Z}_+ \quad (3)$$

$$\text{em que } Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}; e \quad \lambda > 0 \text{ e } \nu \geq 0$$

Propriedades

- ▶ $\frac{P(Y=y-1)}{P(Y=y)} = \frac{y^\nu}{\lambda}$
- ▶ $E(Y) \approx \lambda^{\frac{1}{\nu}} - \frac{\nu-1}{2\nu}$
- ▶ $V(Y) \approx \frac{1}{\nu} E(Y)$

Casos particulares

- ▶ Distribuição Poisson, quando $\nu = 1$
- ▶ Distribuição Bernoulli, quando $\nu \rightarrow \infty$
- ▶ Distribuição Geométrica, quando $\nu = 0, \lambda < 1$

Modelo COM-Poisson

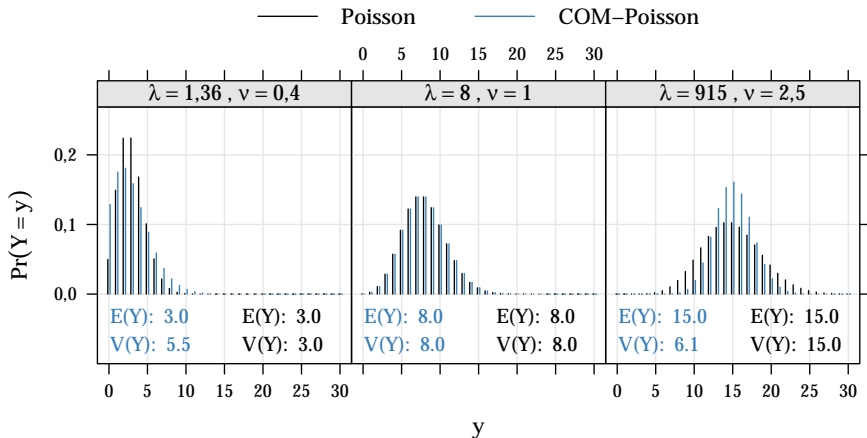


Figura : Probabilidades pela distribuição COM-Poisson para diferentes parâmetros.

Modelo COM-Poisson

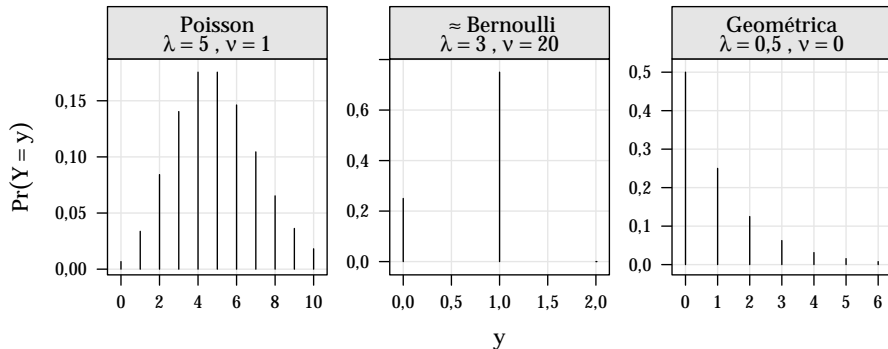


Figura : Exemplos de casos particulares da distribuição COM-Poisson.

1.2

Introdução

Modelos de regressão

Ideia geral

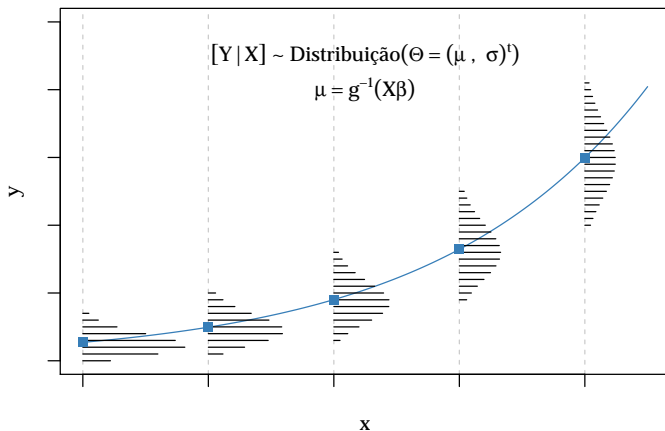


Figura : Exemplificação de um modelo de regressão.

Regressão para dados de Contagem

Sejam Y_1, Y_2, \dots, Y_n variáveis aleatórias condicionalmente independentes, dado o vetor de covariáveis $\underline{x}_i^t = (x_{i1}, x_{i2}, \dots, x_{ip})$;

- ▶ **Regressão log-linear Poisson**

$$Y_i | \underline{x}_i \sim \text{Poisson}(\mu_i)$$

$$\log(\mu_i) = \underline{x}_i^t \beta$$

- ▶ **Regressão Binomial Negativa**

$$Y_i | \underline{x}_i \sim \text{Binomial Negativa}(\mu_i, \theta)$$

$$\log(\mu_i) = \underline{x}_i^t \beta$$

- ▶ **Regressão COM-Poisson**

$$Y_i | \underline{x}_i \sim \text{COM-Poisson}(\lambda_i, \nu)$$

$$g(E(Y_i | \underline{x}_i)) = \log(\lambda_i) = \underline{x}_i^t \beta$$

2

Materiais e Métodos

2.1

Materiais e Métodos
Materiais

Conjuntos de dados

- ▶ 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

- ▶ 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 3.3.1. Principais pacotes:

- ▶ MASS - 7.3.45: ajuste dos modelos binomial negativo;
- ▶ pscl - 1.4.9: ajuste dos modelos para 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.

Relatório do TCC, inteiramente reproduzível:

- ▶ Distribuição \LaTeX - 3.14 (2013/Debian): para editoração do texto;
- ▶ knitr - 1.13: para mesclar códigos R ao texto.
- ▶ Git - 1.9.1: como sistema de versionamento.

2.2

Materiais e Métodos
Métodos

Estimação via máxima verossimilhança

- ➊ Escreva a função de verossimilhança - $\mathcal{L}(\Theta | \underline{y})$
 - ➋ Tome seu logaritmo - $\ell(\Theta | \underline{y})$
 - ➌ As estimativas dos parâmetros são $\hat{\Theta} = \arg \max_{\Theta} \ell(\Theta | \underline{y})$
-
- ▶ Algoritmo IWLS (*Interactive Weighed Leasts Squares*) para os modelos Poisson e Binomial Negativo¹; e
 - ▶ Método *BFGS* para os modelos COM-Poisson.

¹IWLS para os parâmetros de locação e Newton Raphson para o parâmetro de dispersão.

Reparametrização do modelo COM-Poisson

Para garantir o espaço paramétrico do modelo nos reais, faz-se $\phi = \log(\nu)$.

- ▶ $\phi < 0 \Rightarrow$ Superdispersão;
- ▶ $\phi = 0 \Rightarrow$ Equidispersão; e
- ▶ $\phi > 0 \Rightarrow$ Subdispersão

Log-verossimilhança

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

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

Comparação de modelos

- ▶ **log-verossimilhança maximizada**

Maximum log-likelihood

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

- ▶ **CrITÉrio de Informação de Akaike**

Akaike Criterion Information (AIC)

$$AIC = 2(k - \ell(\hat{\Theta}_k, \underline{y}))$$

- ▶ **Teste de razão de verossimilhanças**

Likelihood Ratio Test (LRT)

$$TRV = 2 \left(\ell(\hat{\Theta}_p, \underline{y}) - \ell(\hat{\Theta}_q, \underline{y}) \right) \sim \chi^2_{p-q}$$

- ▶ **Valores preditos para a média**

Confidence Intervals

3

Resultados

3.1

Resultados
Pacote R

Implementação do pacote cmpreg

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

## Regressão (efeitos fixos)
cmp(y ~ preditor, 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)
```

Funções método e conjuntos de dados disponíveis

```
## Datasets
data(package = "cmpreg")

## Principais funções método
summary(model1)           ## Estimativas e testes de Wald
anova(model1, model2)     ## TRV's entre modelos encaixados
cmpptest(model1)          ## TRV para  $H_0: \phi = 0$ 
residuals(model1)         ## Resíduos (ordinários e pearson)
predict(model1)           ## Predição (pontual, intervalar)
```

3.2

Resultados
Caso subdisperso

Capulhos de algodão

Experimento conduzido em casa de vegetação, motivação para (ZEVIANI et al., 2014).

- ▶ Objetivo: avaliar o efeito de desfolha na produção de algodão;
- ▶ Covariáveis experimentais:
 - ▶ Estágio fenológico da planta (est)
(vegetativo, botão floral, florescimento, maça e capulho);
 - ▶ Nível de desfolha aplicada (des)
(0, 25, 50, 75 e 100%).
- ▶ Variável resposta:
 - ▶ Número de capulhos produzidos.
- ▶ Unidade experimental: vaso com duas plantas;
- ▶ Delineamento: Inteiramente casualizado (5 repetições);

Análise descritiva

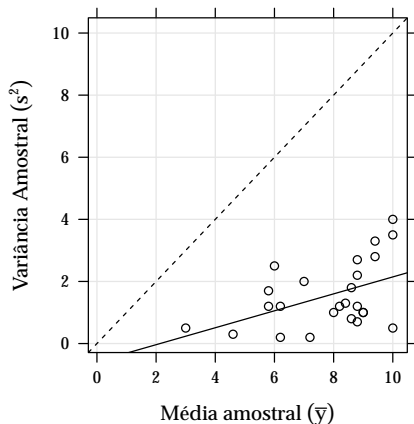
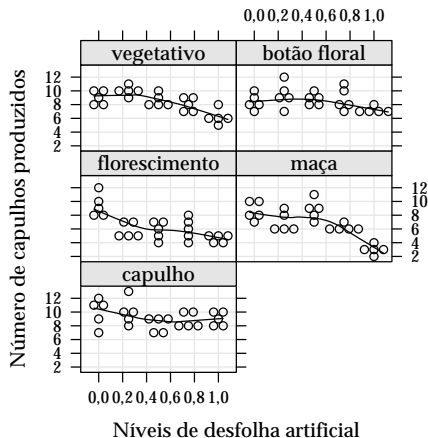


Figura : Disposição das contagens para cada combinação dos fatores (esquerda) e médias e variâncias amostrais (direita).

Modelagem

Preditores considerados:

- ▶ Preditor 1: $g(\mu_{ij}) = \beta_0$
- ▶ Preditor 2: $g(\mu_{ij}) = \beta_0 + \beta_1 \text{des}_i$
- ▶ Preditor 3: $g(\mu_{ij}) = \beta_0 + \beta_1 \text{des}_i + \beta_2 \text{des}_i^2$
- ▶ Preditor 4: $g(\mu_{ij}) = \beta_0 + \beta_{1j} \text{des}_i + \beta_2 \text{des}_i^2$
- ▶ Preditor 5: $g(\mu_{ij}) = \beta_0 + \beta_{1j} \text{des}_i + \beta_{2j} \text{des}_i^2$

j variando nos níveis de estágio fenológico da planta.

Modelos concorrentes:

- ▶ Poisson(μ_{ij})
- ▶ COM-Poisson(λ_{ij}, ϕ)
- ▶ Quasi-Poisson(μ_{ij}, σ^2)

Medidas de ajuste

Tabela : Medidas de ajuste para avaliação e comparação de modelos

Poisson	np	ℓ	AIC	diff np	$P(> \chi^2)$		
Preditor 1	1	-279,93	561,87				
Preditor 2	2	-272,00	548,00	1	0,0001		
Preditor 3	3	-271,35	548,71	1	0,2556		
Preditor 4	7	-258,67	531,35	4	0,0000		
Preditor 5	11	-255,80	533,61	4	0,2193		
COM-Poisson	np	ℓ	AIC	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	1	0,0000	0,794	6,97E-08
Preditor 3	4	-256,09	520,18	1	0,0973	0,816	3,29E-08
Preditor 4	8	-220,20	456,40	4	0,0000	1,392	1,75E-18
Preditor 5	12	-208,25	440,50	4	0,0001	1,585	1,80E-22
Quase-Poisson	np	deviance*	AIC	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		1	0,0000	0,464	5,13E-07
Preditor 3	3	58,36		1	0,0962	0,460	3,66E-07
Preditor 4	7	33,00		4	0,0000	0,278	9,15E-16
Preditor 5	11	27,25		4	0,0002	0,241	3,57E-18

Medidas de ajuste

Tabela : Medidas de ajuste para avaliação e comparação de modelos

Poisson	np	ℓ	AIC	diff np	$P(> \chi^2)$		
Preditor 1	1	-279,93	561,87				
Preditor 2	2	-272,00	548,00	1	0,0001		
Preditor 3	3	-271,35	548,71	1	0,2556		
Preditor 4	7	-258,67	531,35	4	0,0000		
Preditor 5	11	-255,80	533,61	4	0,2193		
COM-Poisson	np	ℓ	AIC	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	1	0,0000	0,794	6,97E-08
Preditor 3	4	-256,09	520,18	1	0,0973	0,816	3,29E-08
Preditor 4	8	-220,20	456,40	4	0,0000	1,392	1,75E-18
Preditor 5	12	-208,25	440,50	4	0,0001	1,585	1,80E-22
Quase-Poisson	np	deviance*	AIC	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		1	0,0000	0,464	5,13E-07
Preditor 3	3	58,36		1	0,0962	0,460	3,66E-07
Preditor 4	7	33,00		4	0,0000	0,278	9,15E-16
Preditor 5	11	27,25		4	0,0002	0,241	3,57E-18

Valores preditos

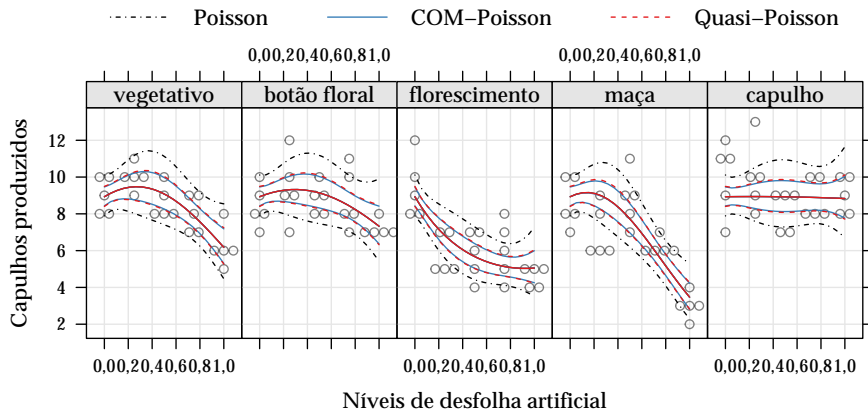


Figura : Curva dos valores preditos com intervalo de confiança de (95%).

3.3

Resultados
Caso superdisperso

Grãos de soja

Experimento conduzido em casa de vegetação, motivação para CITESERAFIM

- ▶ Objetivo: avaliar a produtividade de soja sob solos com diferentes características;
- ▶ Covariáveis experimentais:
 - ▶ Umidade do solo (umid)
(37,5, 62,5 e 62,5%).
 - ▶ Nível de adubação potássica (K)
(0, 30, 60, 120 e 180 mg·dm⁻³);
 - ▶ Indicador de blocagem (bloc)
(I, II, III, VI, V).
- ▶ Variável resposta:
 - ▶ Número de grãos de soja por u.e.
 - ▶ Número de vagens de soja por u.e.
- ▶ Unidade experimental: vaso com duas plantas;
- ▶ Delineamento: Blocos casualizados completos;

Análise descritiva

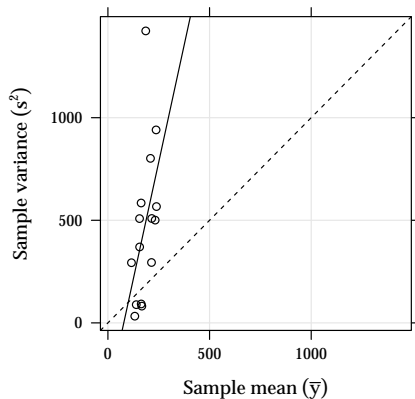
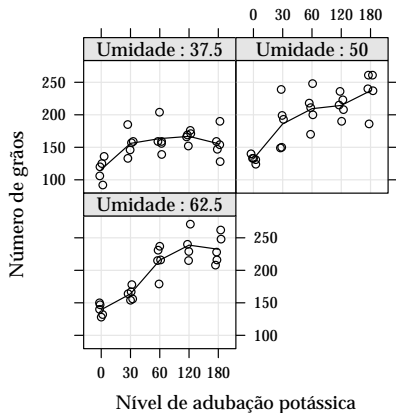


Figura : Disposição do número de grãos nas combinações dos fatores (esquerda) e médias e variâncias amostrais (direita).

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;

γ_j o efeito do j-ésimo nível de umidade aplicado;

δ_k o efeito do k-ésimo nível de adubação potássica; e

α_{jk} o efeito da interação entre o j-ésimo nível de umidade do solo e o k-ésimo nível de adubação potássica.

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 : Medidas de ajuste para avaliação e comparação de modelos

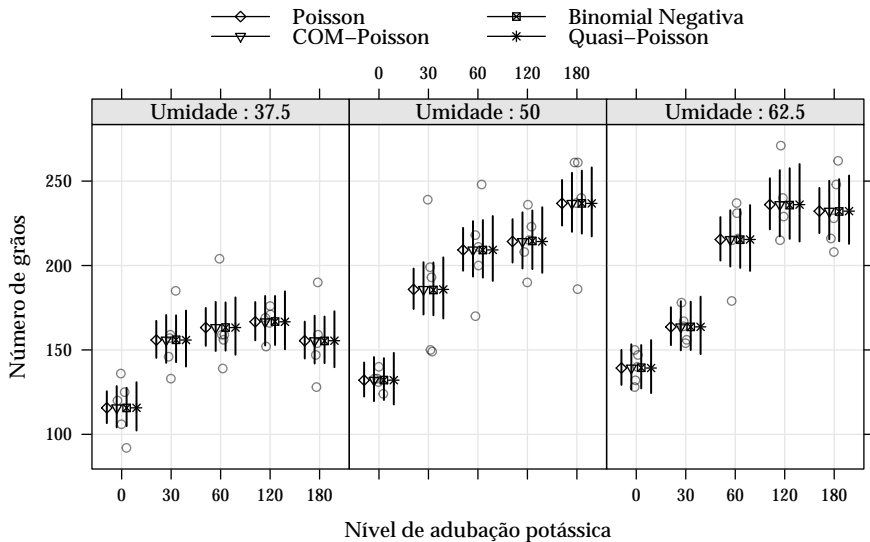
Poisson	np	ℓ	AIC	diff np	$P(> \chi^2)$	
Preditor 1	11	-343,16	708,33			
Preditor 2	19	-321,67	681,34	8	8,83E-07	
COM-Poisson	np	ℓ	AIC	diff np	$P(> \chi^2)$	$\hat{\phi}$
Preditor 1	12	-326,61	677,21			-0,817
Preditor 2	20	-315,64	671,29	8	0,0051	-0,518
Binomial-Negativo	np	deviance*	AIC	diff np	$P(> F)$	$\hat{\sigma}^2$
Preditor 1	12	-326,54	677,07			141,51
Preditor 2	20	-315,39	670,77	8	0,0044	260,94
Quasi-Poisson	np	deviance*	AIC	diff np	$P(> F)$	$\hat{\sigma}^2$
Preditor 1	11	167,71				2,707
Preditor 2	19	124,72		8	0,0300	2,289

Medidas de ajuste

Tabela : Medidas de ajuste para avaliação e comparação de modelos

Poisson	np	ℓ	AIC	diff np	$P(> \chi^2)$	
Preditor 1	11	-343,16	708,33			
Preditor 2	19	-321,67	681,34	8	8,83E-07	
COM-Poisson	np	ℓ	AIC	diff np	$P(> \chi^2)$	$\hat{\phi}$
Preditor 1	12	-326,61	677,21			-0,817
Preditor 2	20	-315,64	671,29	8	0,0051	-0,518
Binomial-Negativo	np	deviance*	AIC	diff np	$P(> F)$	$\hat{\sigma}^2$
Preditor 1	12	-326,54	677,07			141,51
Preditor 2	20	-315,39	670,77	8	0,0044	260,94
Quasi-Poisson	np	deviance*	AIC	diff np	$P(> F)$	$\hat{\sigma}^2$
Preditor 1	11	167,71				2,707
Preditor 2	19	124,72		8	0,0300	2,289

Valores preditos



3.4

Resultados

Caso equidisperso

Vagens de soja

Experimento conduzido em casa de vegetação, motivação para CITESERAFIM

- ▶ Objetivo: avaliar a produtividade de soja sob solos com diferentes características;
- ▶ Covariáveis experimentais:
 - ▶ Umidade do solo (umid)
(37,5, 62,5 e 62,5%).
 - ▶ Nível de adubação potássica (K)
(0, 30, 60, 120 e 180 mg·dm⁻³);
 - ▶ Indicador de blocagem (bloc)
(I, II, III, VI, V).
- ▶ Variável resposta:
 - ▶ Número de grãos de soja por u.e.
 - ▶ Número de vagens de soja por u.e.
- ▶ Unidade experimental: vaso com duas plantas;
- ▶ Delineamento: Blocos casualizados completos;

Análise descritiva

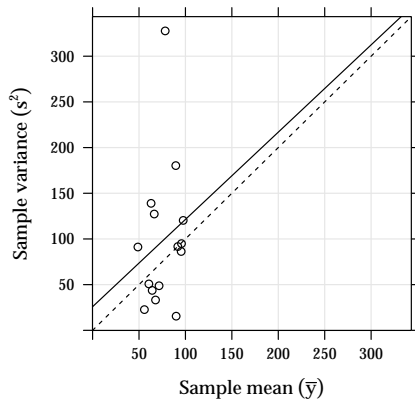
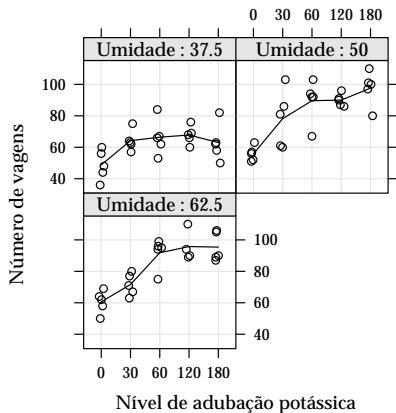


Figura : Disposição do número de grãos nas combinações dos fatores (esquerda) e médias e variâncias amostrais (direita).

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;

γ_j o efeito do j -ésimo nível de umidade aplicado;

δ_k o efeito do k -ésimo nível de adubação potássica; e

α_{jk} o efeito da interação entre o j -ésimo nível de umidade do solo e o k -ésimo nível de adubação potássica.

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 : Medidas de ajuste para avaliação e comparação de modelos

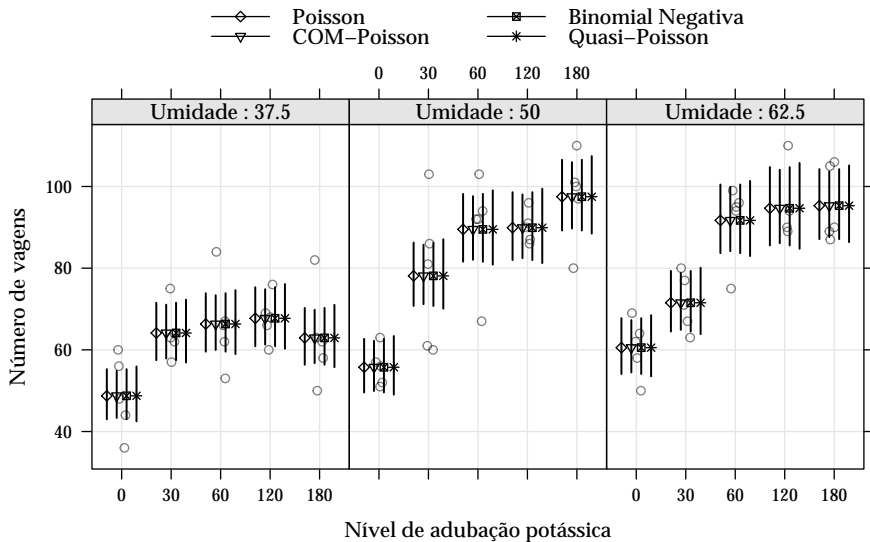
Poisson	np	ℓ	AIC	diff np	$P(> \chi^2)$	
Preditor 1	11	-266,69	555,38			
Preditor 2	19	-259,62	557,23	8	0,0779	
COM-Poisson	np	ℓ	AIC	diff np	$P(> \chi^2)$	$\hat{\phi}$
Preditor 1	12	-266,60	557,20			-0,067
Preditor 2	20	-259,33	558,65	8	0,0685	0,129
Binomial-Negativo	np	ℓ	AIC	diff np	$P(> F)$	$\hat{\theta}$
Preditor 1	12	-266,69	557,37			4,6E+03
Preditor 2	20	-259,62	559,23	8	0,0782	1,0E+06
Quasi-Poisson	np	deviance*	AIC	diff np	$P(> F)$	$\hat{\sigma}^2$
Preditor 1	11	79,43				1,279
Preditor 2	19	65,28		8	0,1875	1,199

Medidas de ajuste

Tabela : Medidas de ajuste para avaliação e comparação de modelos

Poisson	np	ℓ	AIC	diff np	$P(> \chi^2)$	
Preditor 1	11	-266,69	555,38			
Preditor 2	19	-259,62	557,23	8	0,0779	
COM-Poisson	np	ℓ	AIC	diff np	$P(> \chi^2)$	$\hat{\phi}$
Preditor 1	12	-266,60	557,20			-0,067
Preditor 2	20	-259,33	558,65	8	0,0685	0,129
Binomial-Negativo	np	ℓ	AIC	diff np	$P(> F)$	$\hat{\theta}$
Preditor 1	12	-266,69	557,37			4,6E+03
Preditor 2	20	-259,62	559,23	8	0,0782	1,0E+06
Quasi-Poisson	np	deviance*	AIC	diff np	$P(> F)$	$\hat{\sigma}^2$
Preditor 1	11	79,43				1,279
Preditor 2	19	65,28		8	0,1875	1,199

Valores preditos



4

Extensões

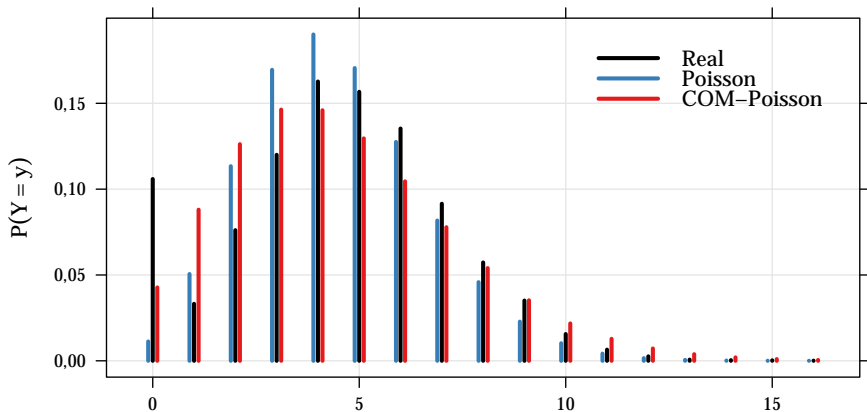
4.1

Extensões

Modelos para excesso de zeros

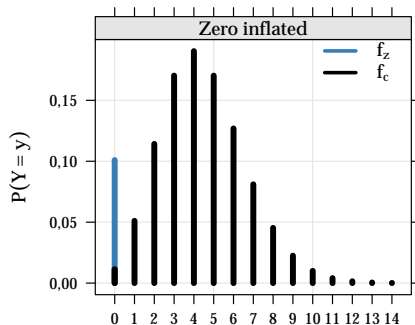
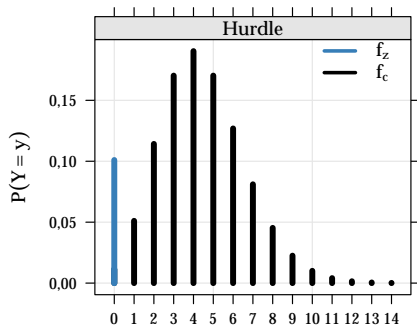
Motivação

Simulação $\mu_{\text{count}} = 2$, $\pi_{\text{zero extra}} = 0,1$



Gerador de dados

- Dois processos compõem a geração dos dados.



Modelo Hurdle COM-Poisson

Função massa de probabilidade de um modelo Hurdle

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

Verossimilhança

$$\mathcal{L}(\phi, \beta, \gamma \mid \underline{y}) = \prod_{i \in \Omega_0} [\pi_i] \prod_{i \in \Omega_+} \left[(1 - \pi_i) \left(\frac{\lambda_i^{y_i}}{(y_i!) e^{\phi} Z(\lambda_i, \phi)} \right) \left(1 - \frac{1}{Z(\lambda_i, \phi)} \right) \right]$$

Sendo

$$\blacktriangleright \Omega_0 = \{i \mid y_i = 0\}$$

$$\blacktriangleright \Omega_+ = \{i \mid y_i > 0\}$$

$$\blacktriangleright \underline{\pi} = \frac{\exp(Z\gamma)}{1 + \exp(Z\gamma)}$$

$$\blacktriangleright \underline{\lambda} = \exp(X\beta)$$

Aplicação

- Estudo observacional com o objetivo de modelar o número de peixes capturados por atividade de pesca esportiva.

Tabela : 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

Aplicação

- Estudo observacional com o objetivo de modelar o número de peixes capturados por atividade de pesca esportiva.

Tabela : 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

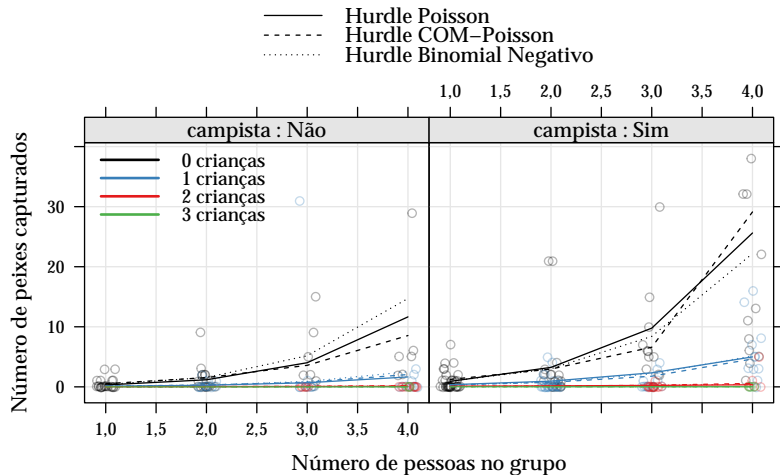


Figura : Valores preditos do número de peixes capturados.

4.2

Extensões

Modelos de efeitos aleatórios

Motivação

- Correlação entre grupos de indivíduos induzida pelo delineamento experimental ou estrutura do problema.

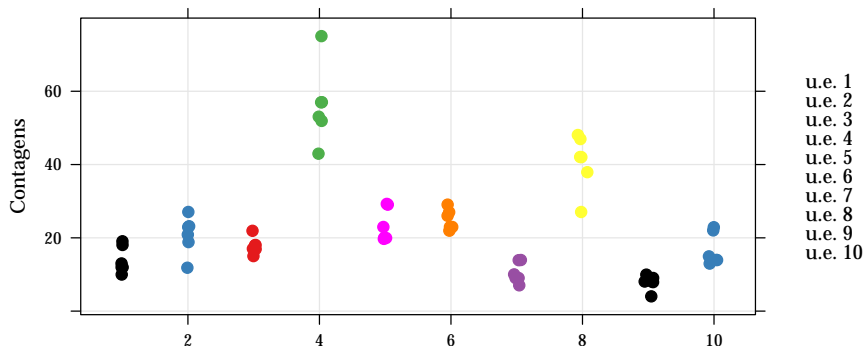


Figura : Contagens que apresentam um efeito aleatório da unidade experimental (u.e.)

Modelo COM-Poisson Misto

Estrutura hierárquica do modelo

$$Y | \mathbf{b} \sim f_*(\mu, \phi)$$

$$g(\mu) = \beta_0 + \mathbf{b}_i$$

$$\mathbf{b}_i \sim N(\mathbf{0}, \sigma^2)$$

Verossimilhança

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

Aplicação

Tabela : 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

Aplicação

Tabela : 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

Valores preditos

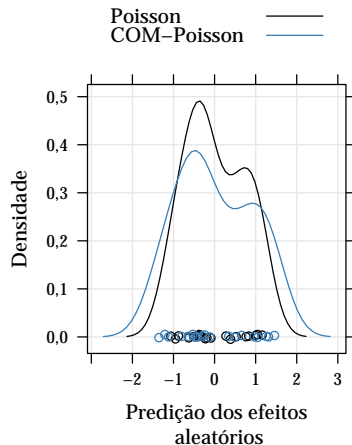
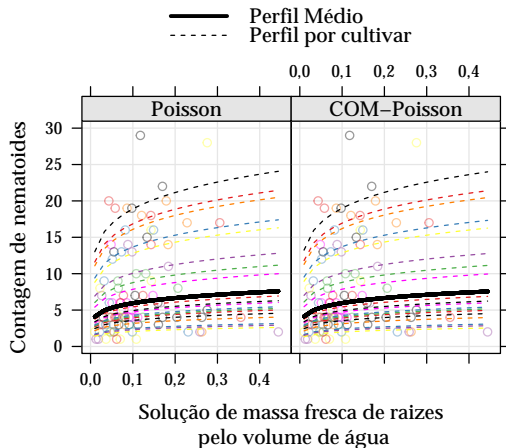


Figura : Valores preditos nos modelos de efeitos mistos.

5

Conclusões

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.

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.

Trabalhos futuros

Sugestões para continuidade da pesquisa:

- ▶ 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
- ▶ Expandir o modelo COM-Poisson de efeitos aleatórios:

Publicização



cmpreg: Implementação computacional em formato de pacote R.








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



monografia: Redação do relatório e scripts de análise. Tudo reproduzível.

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

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.
-  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.
-  PAULA, G. A. *Modelos de regressão com apoio computacional*. IME-USP São Paulo, 2013. Disponível em: <https://www.ime.usp.br/~giapaula/textoregressao.h>.
-  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>.
-  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.
-  WINKELMANN, R.; ZIMMERMANN, K. F. *Count data models for demographic data*. 1994. 205–221, 223 p.
-  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>.