

# Modelo de Regressão COM-Poisson: *defoliation*

Eduardo E. R. Junior & Walmes M. Zeviani

Esta *vignette*<sup>1</sup> faz parte do trabalho de conclusão de curso em Estatística na Universidade Federal do Paraná que objetiva a exploração de modelos de regressão para variáveis de contagem, por Eduardo Junior sob orientação de Walmes Zeviani.

## Conjunto de dados

O conjunto de dados<sup>2</sup> contém 125 observações provenientes de um experimento em casa de vegetação inteiramente casualizado com 5 repetições, cujas plantas de algodão (*Gossypium hirsutum*) foram submetidas à níveis de desfolha artificial (5 níveis) combinados com o estágio fenológico da planta na aplicação (5 níveis). As variáveis presentes no conjunto de dados são:

- phenol: estágio fenológico durante a aplicação da desfolha ( *vegetative*, *flower bud*, *blossom*, *boll*, *boll open*);
- defol: nível de desfolha artificial aplicado ( 0, 0.25, 0.5, 0.75, 1);
- bolls: número de capulhos produzidos ao final do ciclo cultura

Pela Figura ?? notamos a relevância das covariáveis do estudo, pois as curvas de suavização são razoavelmente distintas para cada estágio (phenol) e há uma tendência evidente ao considerar a variação do nível de desfolha (defol) aplicado, em (a) e (b). Além disso, de forma complementar o gráfico em (c) apresenta as médias e variâncias amostrais calculadas nas repetições de cada tratamento e notamos a clara evidência de subdispersão, pois todos os pontos estão abaixo da linha 1 pra 1 que representa a equidispersão.

<sup>1</sup> código fonte disponível em <https://gitlab.c3sl.ufpr.br/eej12/tcc-eduardo>

<sup>2</sup> Veja detalhes em: **Impacto de diferentes níveis de desfolha artificial nos estádios fenológicos do algodoeiro.**

<http://www.scielo.mec.pt/pdf/rca/v35n1/v35n1a16.pdf>

Visualizando os dados no R.

```
## git <- paste0(
##   "http://git.leg.ufpr.br/",
##   "leg/legTools.git")
## devtools::install_git(git)

library(legTools); data(defoliation)

summary(defoliation[, c(1, 2)])

      phenol      defol
vegetative:25   Min.   :0.00
flower bud:25   1st Qu.:0.25
blossom :25     Median :0.50
boll     :25     Mean   :0.50
boll open :25    3rd Qu.:0.75
                        Max.   :1.00

summary(defoliation[, -c(1, 2)])

      rept      bolls
Min.   :1   Min.   : 2.000
1st Qu.:2   1st Qu.: 7.000
Median :3   Median : 8.000
Mean   :3   Mean   : 7.824
3rd Qu.:4   3rd Qu.: 9.000
Max.   :5   Max.   :13.000
```

Encurtando nomes, para facilitar a escrita dos códigos

```
## Estreitando nomes
defol <- defoliation
(names(defol) <- substr(
  names(defoliation), 1, 3))

[1] "phe" "def" "rep" "bol"
```

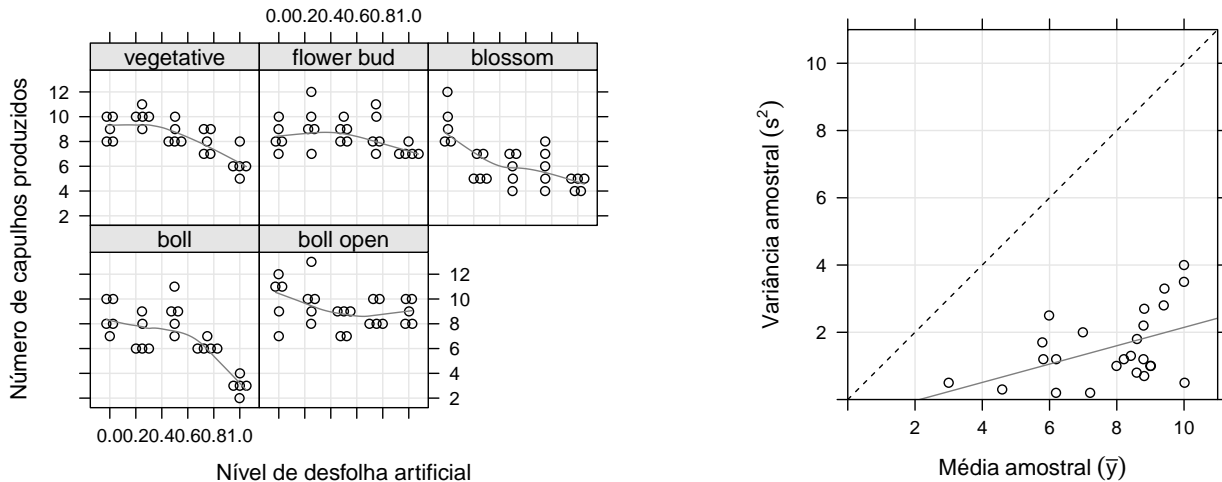


Figura 1: (a) Número de capulhos produzidos pelo nível de desfolha estratificado por estágio da planta. (b) Variâncias em funções das médias amostrais calculadas com base nas 5 repetições de cada tratamento

### Modelos propostos

Nesta seção ajustamos modelos estatísticos com 3 diferentes distribuições marginais,  $[Y | X]$  associadas, *Poisson*, *Gamma Count*<sup>3</sup> e *COM-Poisson*. Para comparação considerou-se o ajuste do modelo *quasi-Poisson*. Além disso diferentes estruturas para o preditor linear foram consideradas:

<sup>3</sup> As análises foram baseadas em **The Gamma-count distribution in the analysis of experimental underdispersed data**  
<http://www.leg.ufpr.br/~walmes/papercompanions/gammacount2014/papercomp.html>

- Modelo 0:  $\beta_0$  (1 parâmetro)
- Modelo 1:  $\beta_0 + \beta_1 def$  (2 parâmetros)
- Modelo 2:  $\beta_0 + \beta_1 def + \beta_2 def^2$  (3 parâmetros)
- Modelo 3:  $\beta_0 + \beta_{1phe} def + \beta_2 def^2$  (7 parâmetros)
- Modelo 4:  $\beta_0 + \beta_{1phe} def + \beta_{2phe} def^2$  (11 parâmetros)

```
## -----
## Estimação dos modelos
## -----
## Modelo Poisson
cpP0 <- glm(bol ~ 1, data = defol, family = poisson)
cpP1 <- glm(bol ~ def, data = defol, family = poisson)
cpP2 <- glm(bol ~ def + I(def^2), data = defol, family = poisson)
cpP3 <- glm(bol ~ phe: def + I(def^2), data = defol, family = poisson)
cpP4 <- glm(bol ~ phe:(def + I(def^2)), data = defol, family = poisson)
## -----
## Modelo Quase Poisson
cpQ0 <- glm(formula(cpP0), data = defol, family = quasipoisson)
cpQ1 <- glm(formula(cpP1), data = defol, family = quasipoisson)
cpQ2 <- glm(formula(cpP2), data = defol, family = quasipoisson)
cpQ3 <- glm(formula(cpP3), data = defol, family = quasipoisson)
cpQ4 <- glm(formula(cpP4), data = defol, family = quasipoisson)
## -----
## Modelo Contagem Gama
cpG0 <- poi2cg(cpP0)
cpG1 <- poi2cg(cpP1)
cpG2 <- poi2cg(cpP2)
cpG3 <- poi2cg(cpP3)
```

```
cpG4 <- poi2cg(cpP4)

##-----
## Modelo COM-poisson
cpC0 <- glm.comp(formula(cpP0), data = defol)
cpC1 <- glm.comp(formula(cpP1), data = defol)
cpC2 <- glm.comp(formula(cpP2), data = defol)
cpC3 <- glm.comp(formula(cpP3), data = defol)
cpC4 <- glm.comp(formula(cpP4), data = defol)
```

```
anova(cpP0, cpP1, cpP2, cpP3, cpP4, test = "Chisq")

Analysis of Deviance Table

Model 1: bol ~ 1
Model 2: bol ~ def
Model 3: bol ~ def + I(def^2)
Model 4: bol ~ phe:def + I(def^2)
Model 5: bol ~ phe:(def + I(def^2))
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      124      75.514
2      123      59.650  1  15.8643 6.805e-05 ***
3      122      58.357  1   1.2926  0.2556
4      118      32.997  4  25.3604 4.258e-05 ***
5      114      27.255  4   5.7420  0.2193
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(cpQ0, cpQ1, cpQ2, cpQ3, cpQ4, test = "F")

Analysis of Deviance Table

Model 1: bol ~ 1
Model 2: bol ~ def
Model 3: bol ~ def + I(def^2)
Model 4: bol ~ phe:def + I(def^2)
Model 5: bol ~ phe:(def + I(def^2))
  Resid. Df Resid. Dev Df Deviance      F      Pr(>F)
1      124      75.514
2      123      59.650  1  15.8643 65.8202 6.343e-13 ***
3      122      58.357  1   1.2926  5.3630 0.0223564 *
4      118      32.997  4  25.3604 26.3047 1.846e-15 ***
5      114      27.255  4   5.7420  5.9558 0.0002176 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova.cg(cpG0, cpG1, cpG2, cpG3, cpG4)

      ll npar two.ll.dif npar.dif      pvalue
1 -272.3961    2      NA      NA      NA
2 -257.3503    3 30.091519    1 4.121296e-08
3 -255.9811    4  2.738461    1 9.795906e-02
4 -220.1455    8 71.671185    4 1.007088e-14
5 -208.3865   12 23.518023    4 9.975695e-05
```

```
anova.cmp(cpC0, cpC1, cpC2, cpC3, cpC4)

      ll npar two.ll.dif npar.dif      pvalue
1 -272.4794    2      NA      NA      NA
2 -257.4636    3 30.031524    1 4.250794e-08
3 -256.0893    4  2.748565    1 9.734174e-02
4 -222.5869    8 67.004730    4 9.726501e-14
5 -214.4700   12 16.233801    4 2.720960e-03
```

Tabela 1: Log-verossimilhanças

Modelos	Gamma.Count	COM.Poisson
0	-272.40	-272.48
1	-257.35	-257.46
2	-255.98	-256.09
3	-220.15	-222.59
4	-208.39	-214.47