

Modelos Lineares Generaliza

Princípios e variável contagem

Alexandre Adalardo de Oliveira

PlanECO 2025

GLM: Princípios

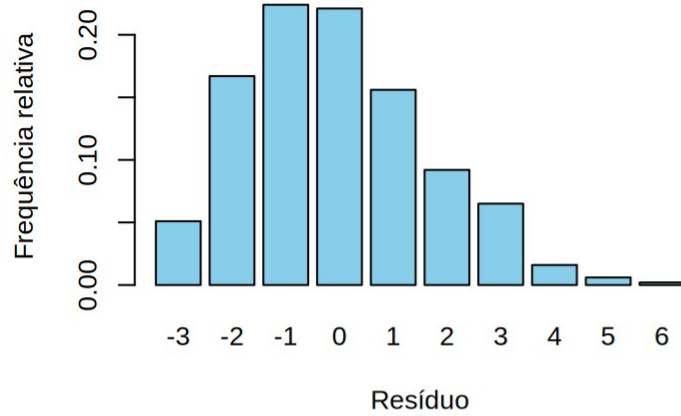
Modelos Lineares Generalizados (GLMs)

Conceitos Tratados

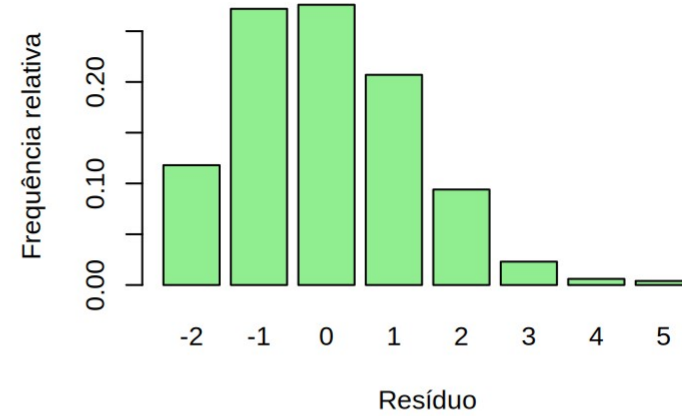
- estrutura dos resíduos não Gaussiano
- função de ligação
- preditora linear
- inverso da função de ligação
- **conceitos importantes associados**
 - máxima verossimilhança (maximum Likelihood)
 - deviance (“desvio”)

Distribuição Probabilísticas não Gaussiana

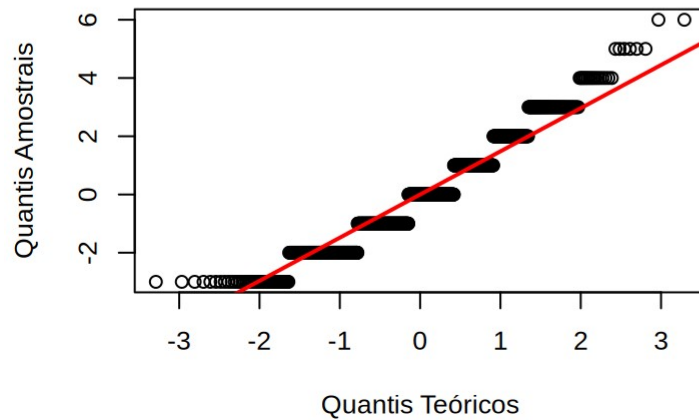
Resíduos - Poisson (lambda = 3)



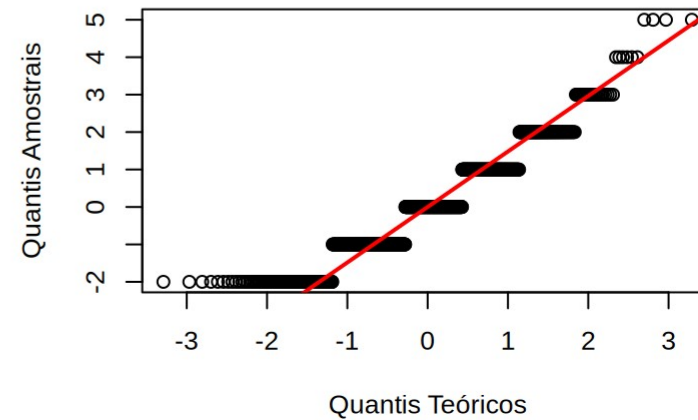
Resíduos - Binomial (n= 10, p = 0.2)



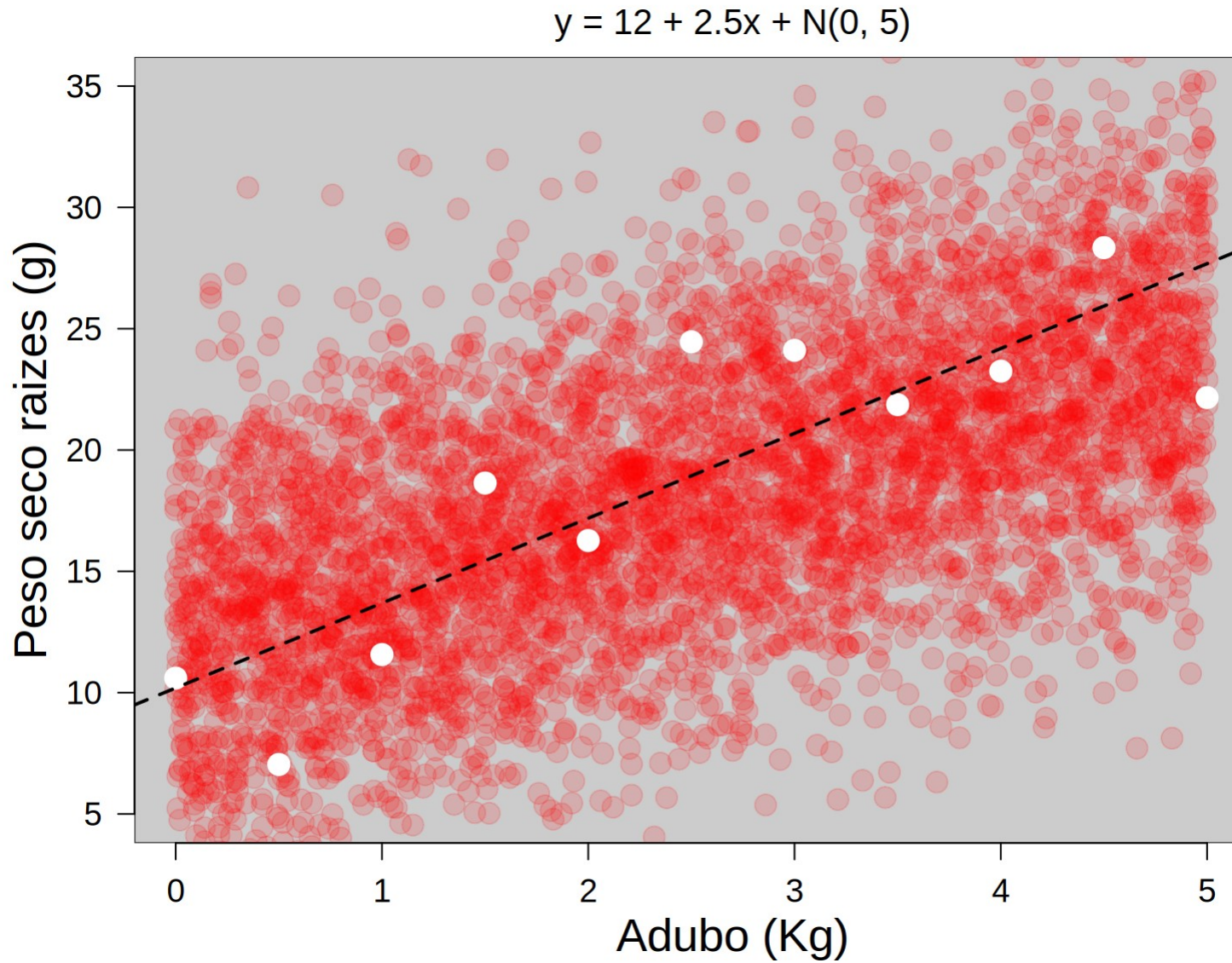
Q-Q Plot - Resíduos Poisson



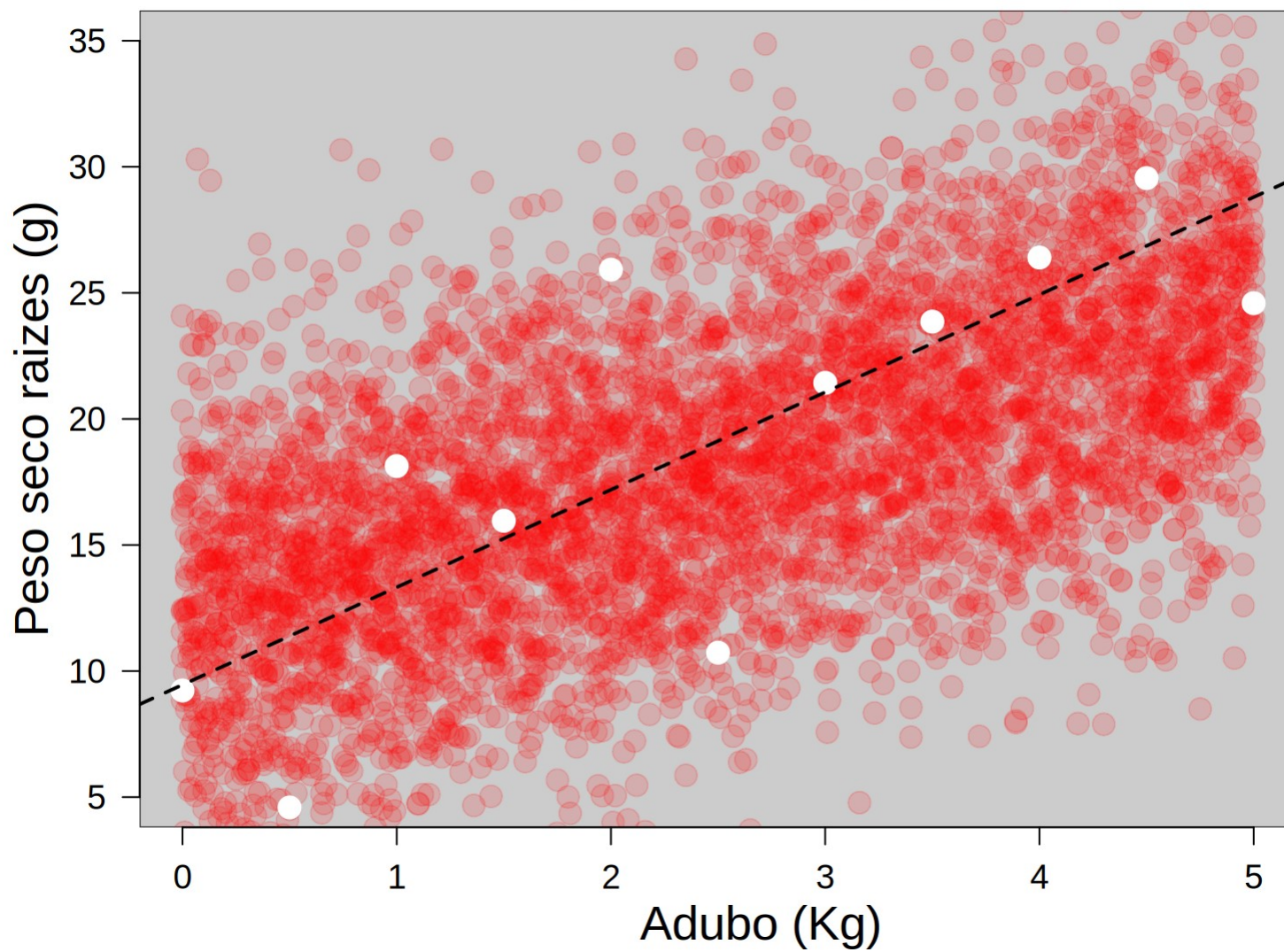
Q-Q Plot - Resíduos Binomial



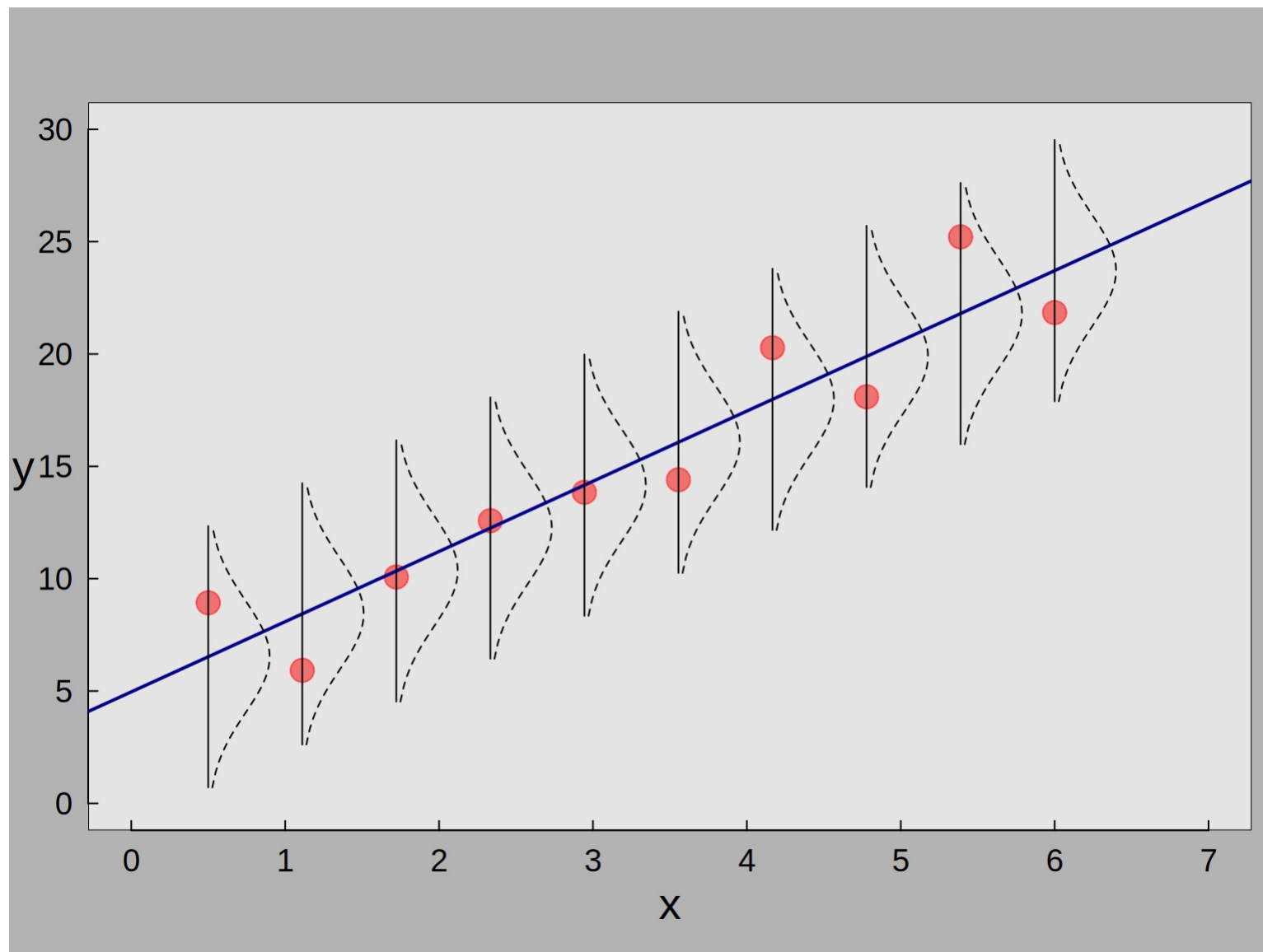
Verossimilhança



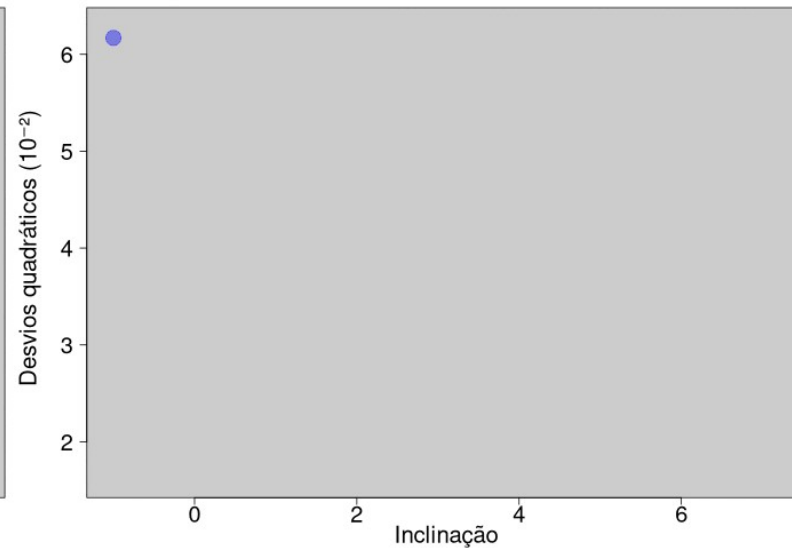
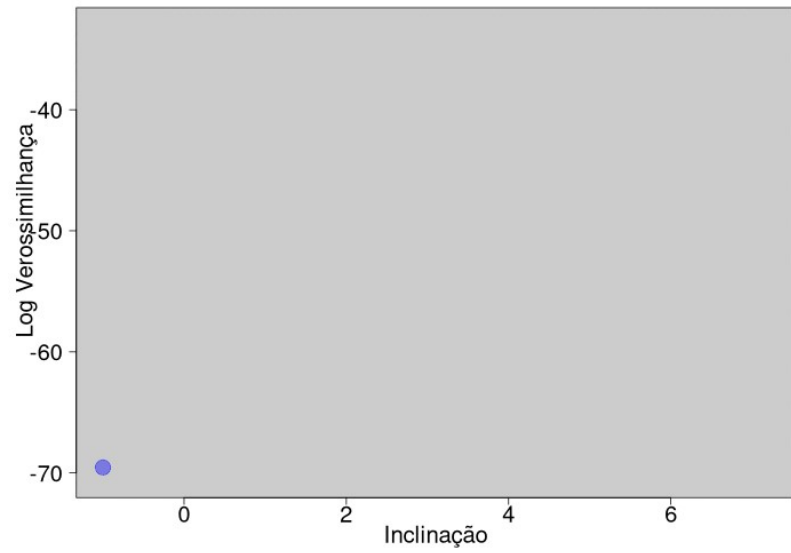
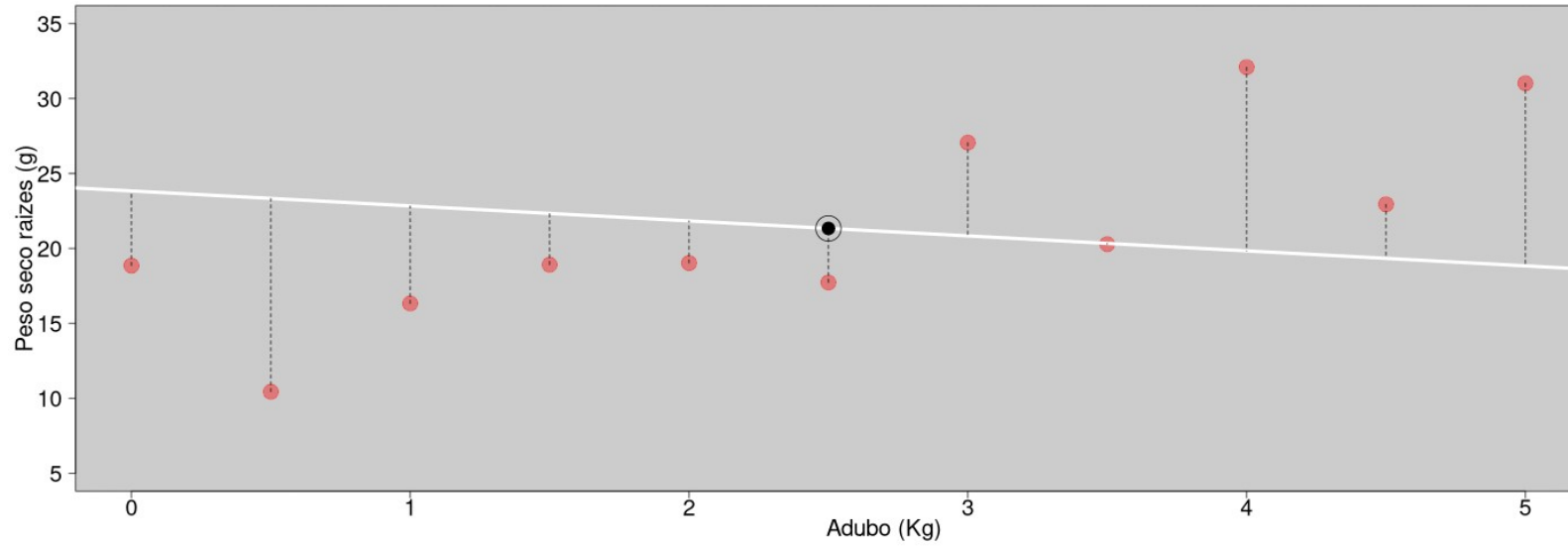
$$y = 12 + 2.5x + N(0, 5)$$



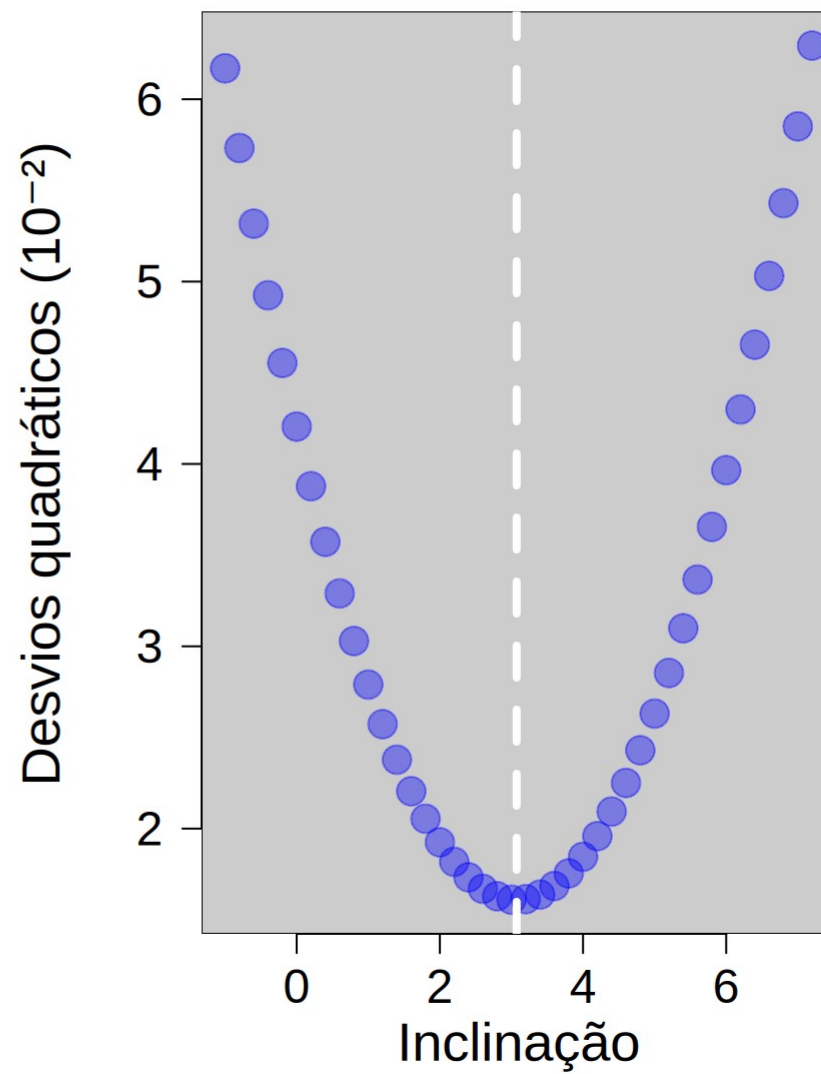
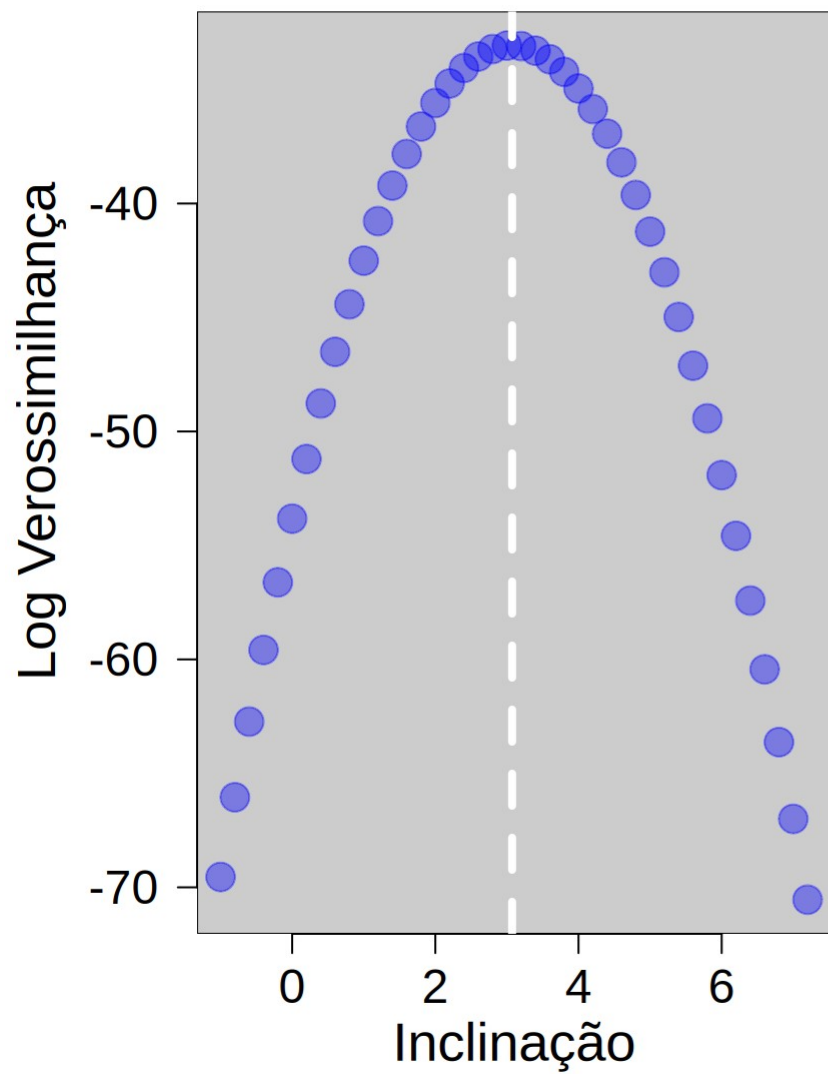
Probabilidade de gerar os dados



Verossimilhança



Máxima Verossimilhança



Deviance

$$Dev = 2 * \frac{\mathcal{L}_{sat}}{\mathcal{L}_{mod}}$$

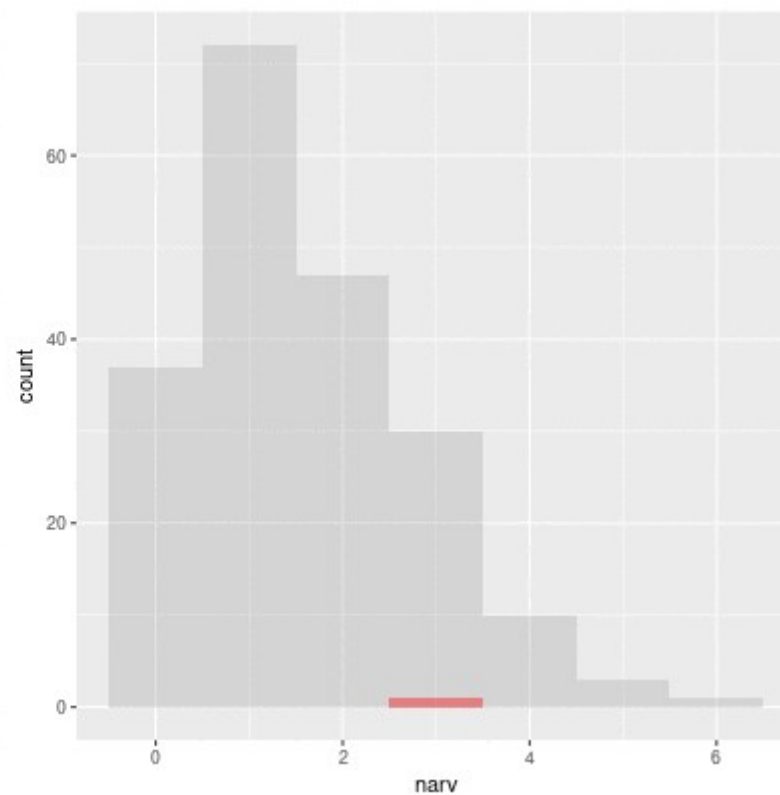
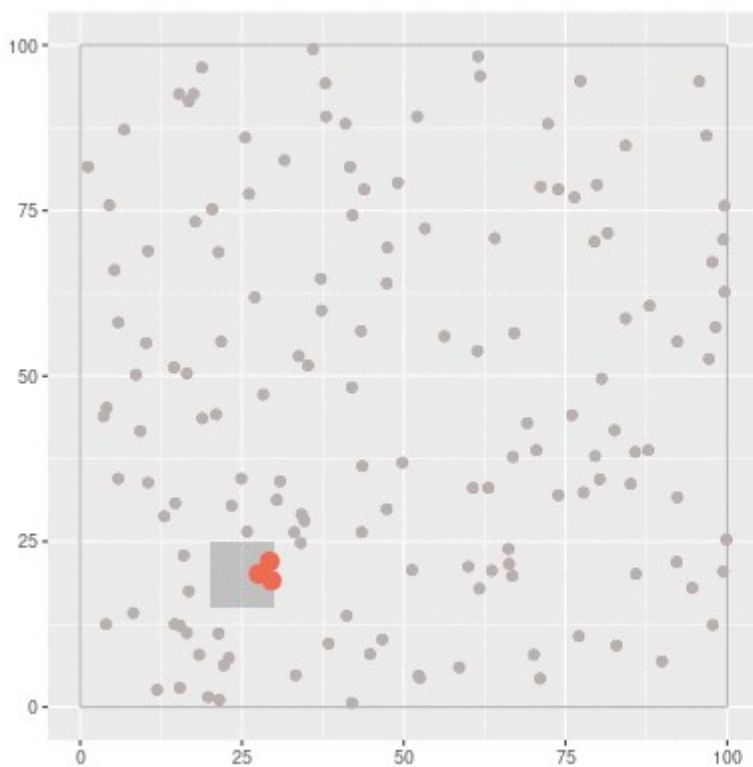
$$Dev = 2 * (\log(\mathcal{L}_{sat}) - \log(\mathcal{L}_{mod}))$$

Family (error structure)	Deviance	Variance function
gaussian	$\sum (y - \bar{y})^2$	1
poisson	$2 \sum y \ln(y/\mu) - (y - \mu)$	μ
binomial	$2 \sum y \ln(y/\mu) + (n - y) \ln(n - y)/(n - \mu)$	$\frac{\mu(n - \mu)}{n}$
Gamma	$2 \sum (y - \mu)/y - \ln(y/\mu)$	μ^2
inverse.gaussian	$\sum (y - \mu)^2 / (\mu^2 y)$	μ^3

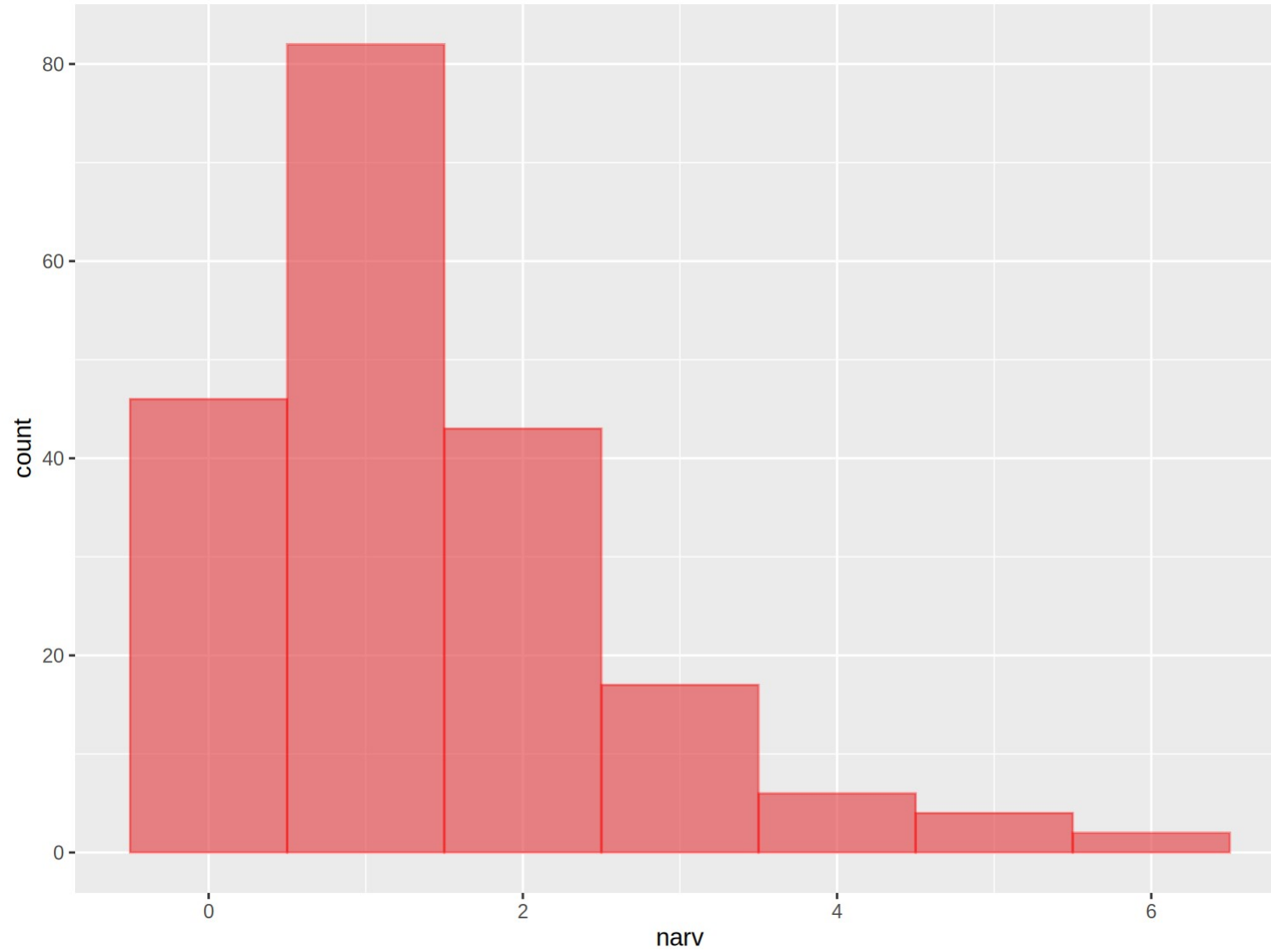
Estrutura dos resíduos

Completa Aleatoriedade Espacial

- Contagem de árvores em Parcelas



Dados de Contagem



Dados de Contagem

- contagem é limitada no zero
- valores discretos (inteiros)
- variância não é constante (aumenta com a média)
- resíduo não tem distribuição normal
- relação não linear com preditoras contínuas

Linearizável
Familia Exponencial

GLMs: linearização

Função de ligação

- aplicada na ~~Esperança~~ ~~Esperança~~ resposta:

$$\eta = g(\mu)$$

Preditor Linear

- o preditor ~~é~~ a resposta linear de uma ou + preditoras

$$\eta = \alpha + \sum_{j=1}^p \beta_j x_{ij}$$

Função inversa

- retorna à escala da preditora

$$Y = g(\eta)^{-1}$$

Função de ligação: canônica

- dependendo da estrutura dos resíduos há uma função de ligação padrão

resposta	resíduos	ligação
contínua	gaussiano	identidade
contagem	poisson	log
proporção	binomial	logit
binária	binomial	logit

LMs x GLMs

Função de ligação identidade

$$\eta = I(\hat{Y})$$

$$\eta = \hat{Y}$$

Estrutura do resíduo normal

$$\epsilon = N(0, \sigma)$$

GLMs x Tranformar LMs

Modelo Transformado de LMs

$$f(y)$$

$$f(\hat{Y} + \epsilon)$$

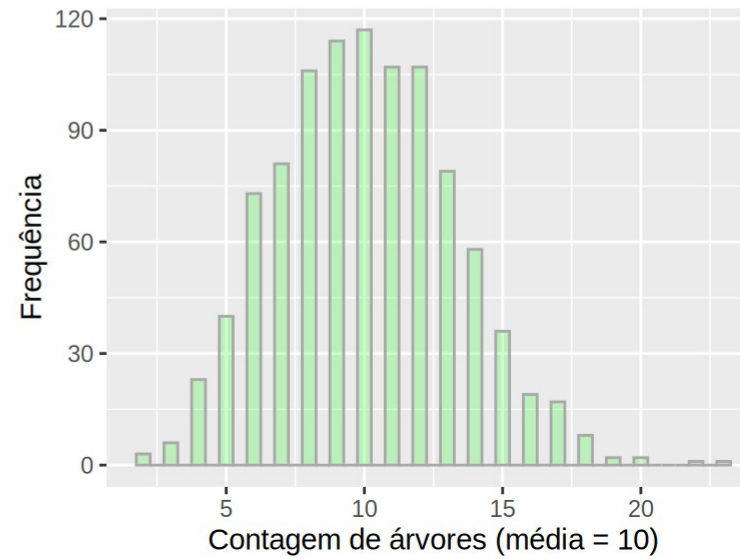
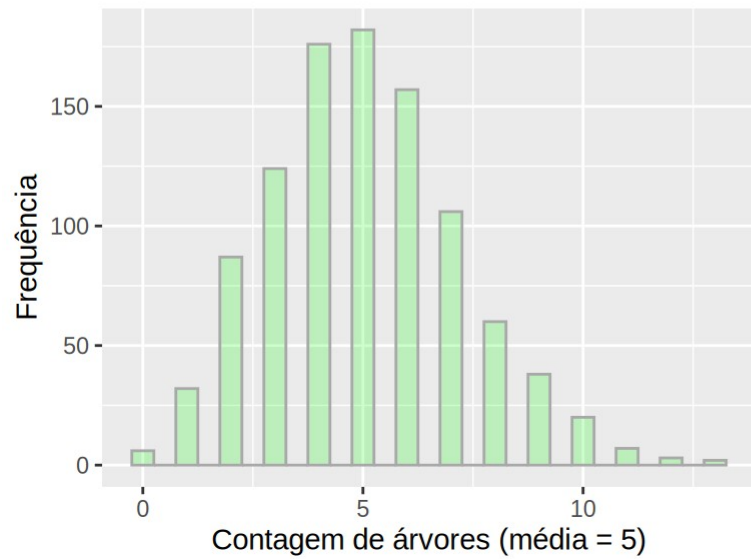
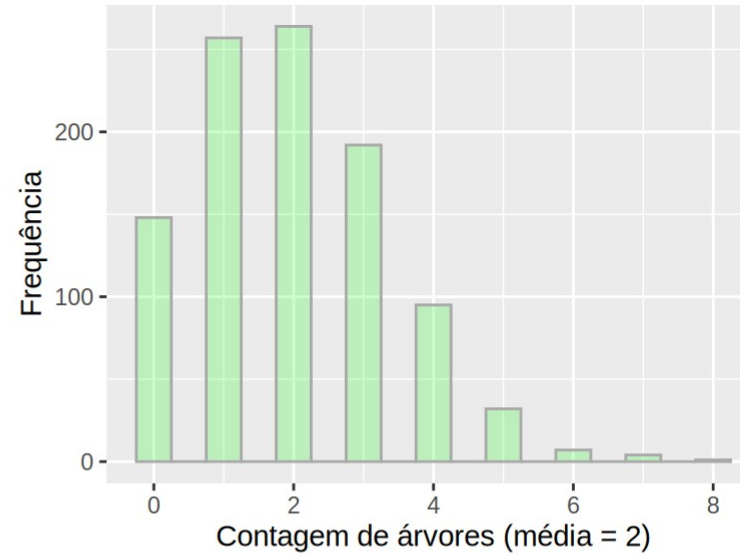
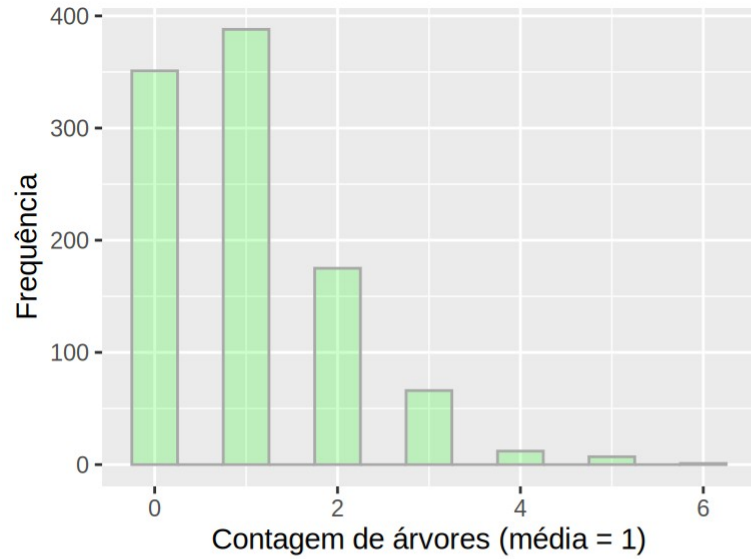
- transforma o valor observado (predito + resíduo)

Função de Ligação

$$f(\hat{Y})$$

- tranforma apenas a experança da variável aleatória
- preserva a estrutura do resíduo
- modela a estrutura do resíduo na escala original (Poisson, Binomial, Beta, Gamma...)

Estruturas dos resíduos



GLM: contagem

Dados de Contagem

- contagem é limitada no zero
- valores discretos (inteiros)
- variância não é constante (aumenta com a média)
- resíduos não tem distribuição normal

Função de ligação

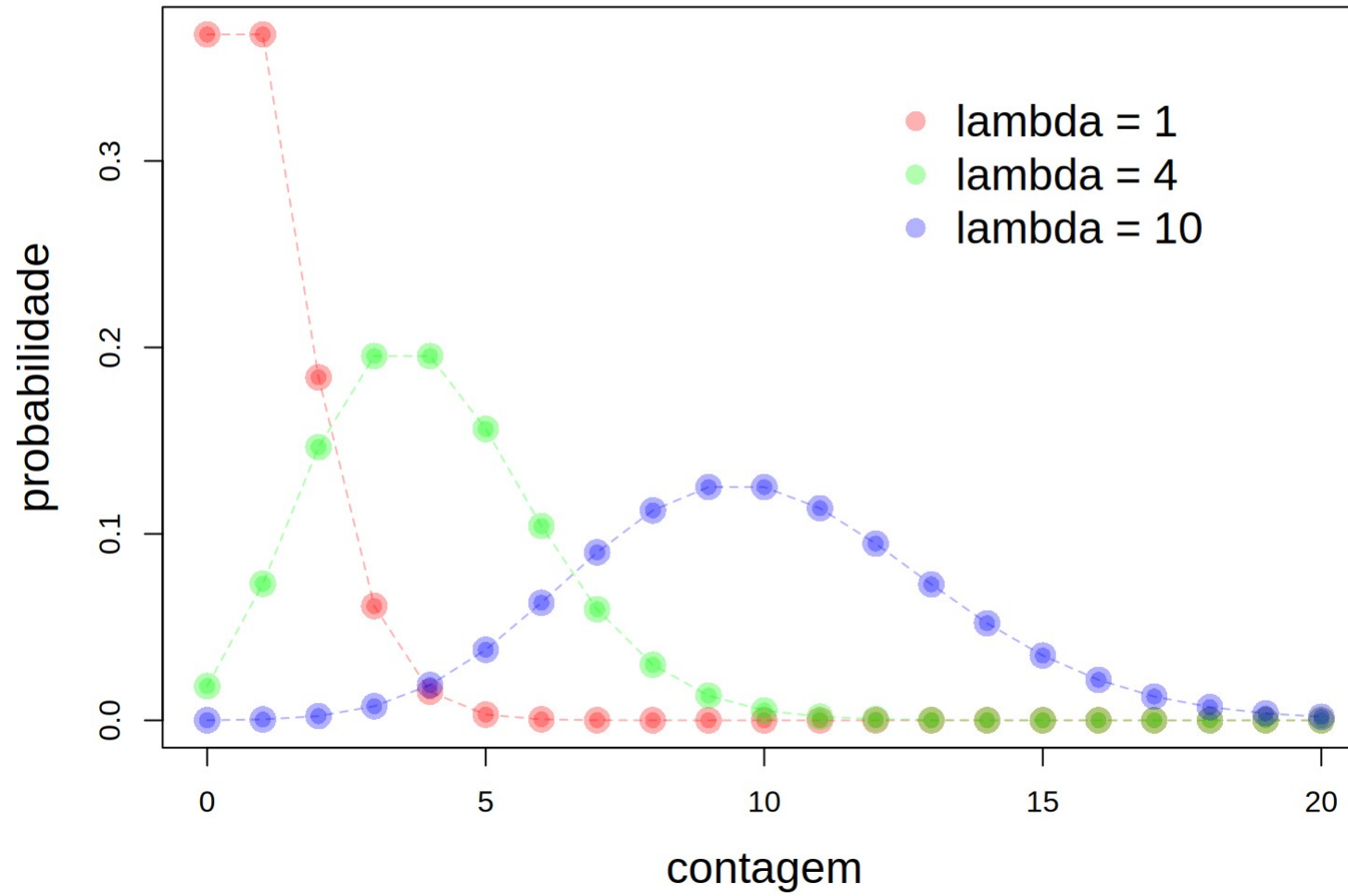
$$\log(\alpha + \sum \beta_i x_i)$$

Estrutura dos resíduos

- **Poisson**

Poisson

Distribuição de Probabilidade Poisson

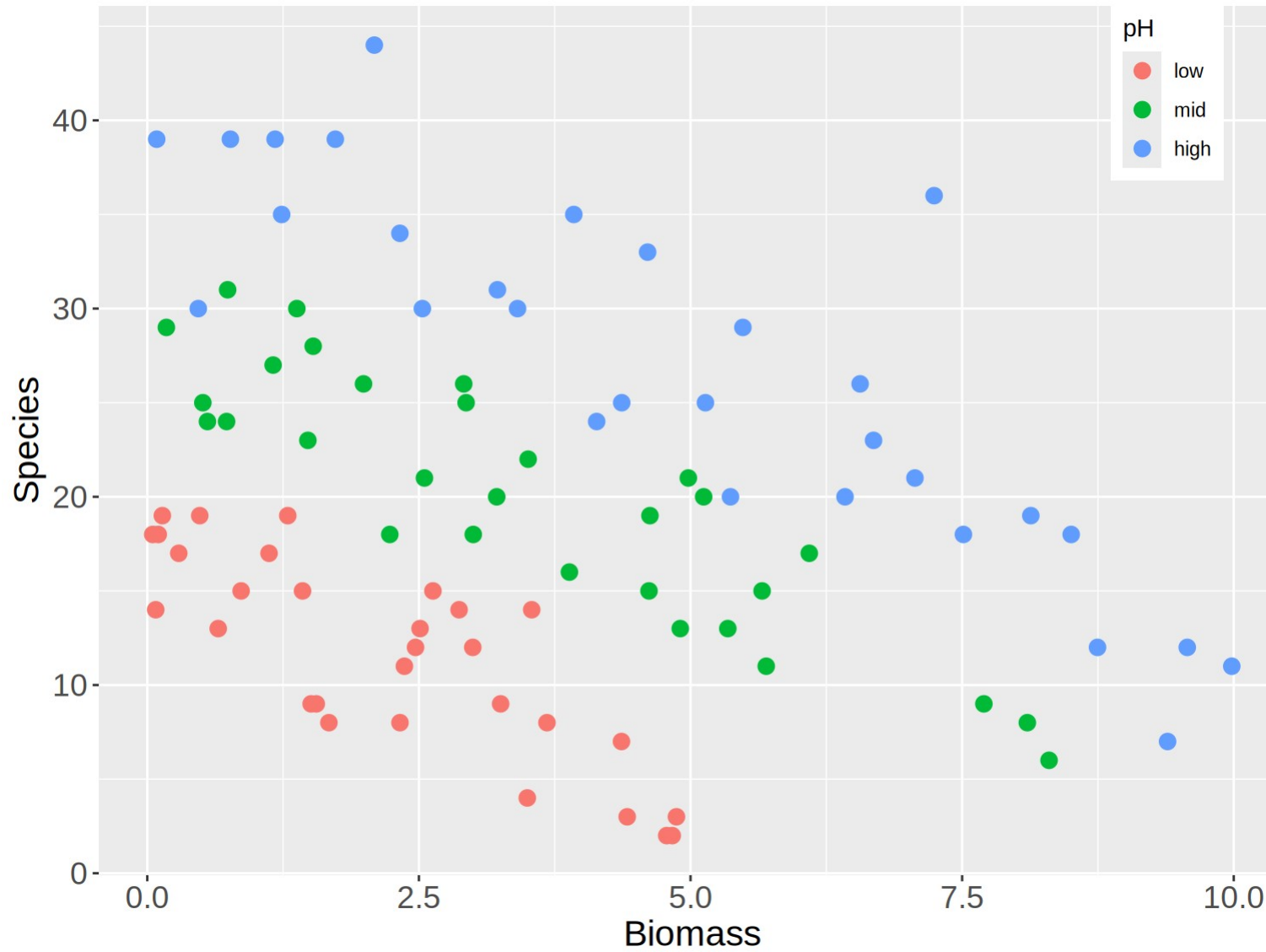


Exemplo: Contagem de espécies

- Biomassa e ph do solo estão relacionados à coexistência de espécies? (n = 90)

Species	Biomass	pH
30	0.4692972	high
39	1.7308704	high
44	2.0897785	high
29	0.1757627	mid
30	1.3767783	mid
21	2.5510426	mid
18	0.1008479	low
19	0.1385961	low
15	0.8635151	low

Gráfico: riqueza de espécies



Modelo

```
glm01 <- glm(Species ~ Biomass + pH + Biomass:pH, family =  
  poisson, data= arv)  
summary(glm01)
```

```
##  
## Call:  
## glm(formula = Species ~ Biomass + pH + Biomass:pH, family  
= poisson,  
## data = arv)  
##  
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 2.95255 0.08240 35.833 < 2e-16 ***  
## Biomass -0.26216 0.03803 -6.893 5.47e-12 ***  
## pHmid 0.48411 0.10723 4.515 6.34e-06 ***  
## pHhigh 0.81557 0.10284 7.931 2.18e-15 ***  
## Biomass:pHmid 0.12314 0.04270 2.884 0.003927 **  
## Biomass:pHhigh 0.15503 0.04003 3.873 0.000108 ***  
## ---  
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '  
' 1  
##  
## (Dispersion parameter for poisson family taken to be 1)  
##  
## Null deviance: 452.346 on 89 degrees of freedom  
## Residual deviance: 83.201 on 84 degrees of freedom  
## AIC: 514.39
```

```
##  
## Number of Fisher Scoring iterations: 4
```

Simplificando o Modelo

Modelos Poisson: Qui-quadrado

```
glm02 <- glm(Species ~ Biomass + pH, family = poisson, data= arv)
anova(glm02, glm01, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: Species ~ Biomass + pH
## Model 2: Species ~ Biomass + pH + Biomass:pH
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         86      99.242
## 2         84      83.201  2    16.04 0.0003288 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anova do modelo

Comparação sequencial

```
anova(glm01, test = "Chisq")
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model: poisson, link: log
```

```
##
```

```
## Response: Species
```

```
##
```

```
## Terms added sequentially (first to last)
```

```
##
```

```
##
```

```
##           Df Deviance Resid. Df Resid. Dev   Pr(>Chi)
```

```
## NULL                89      452.35
```

```
## Biomass             1       44.673    88      407.67 2.328e-11 ***
```

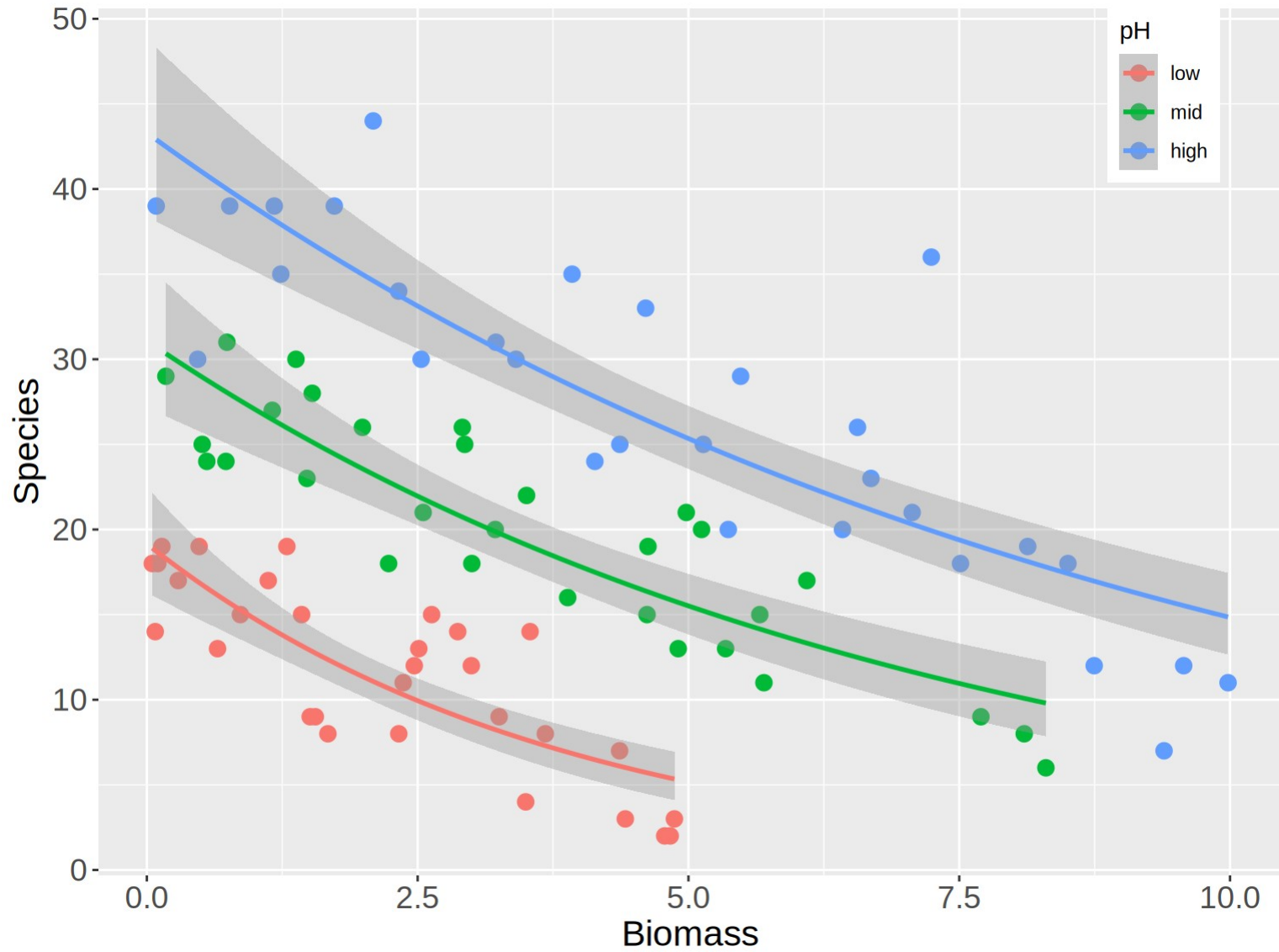
```
## pH                  2      308.431    86        99.24 < 2.2e-16 ***
```

```
## Biomass:pH         2       16.040    84        83.20 0.0003288 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Gráfico do nosso modelo



Mínimo adecuado

```
##
## Call:
## glm(formula = Species ~ Biomass + pH + Biomass:pH, family
## = poisson,
## data = arv)
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) 2.95255 0.08240 35.833 < 2e-16 ***
## Biomass -0.26216 0.03803 -6.893 5.47e-12 ***
## pHmid 0.48411 0.10723 4.515 6.34e-06 ***
## pHhigh 0.81557 0.10284 7.931 2.18e-15 ***
## Biomass:pHmid 0.12314 0.04270 2.884 0.003927 **
## Biomass:pHhigh 0.15503 0.04003 3.873 0.000108 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 452.346 on 89 degrees of freedom
## Residual deviance: 83.201 on 84 degrees of freedom
## AIC: 514.39
##
## Number of Fisher Scoring iterations: 4
```

$$2.95 - 0.26bio + 0.48ph_{mid} + 0.82ph_{high}$$

Predito pelo modelo

$$2.95 - 0.26bio + 0.48ph_{mid} + 0.82ph_{high}$$

- **pH baixo, biomassa 0.47**

```
## (Intercept)      Biomass
##      2.952552    -0.262158
```

$$2.95 - 0.26 * 0.47$$

```
## [1] 2.829338
```

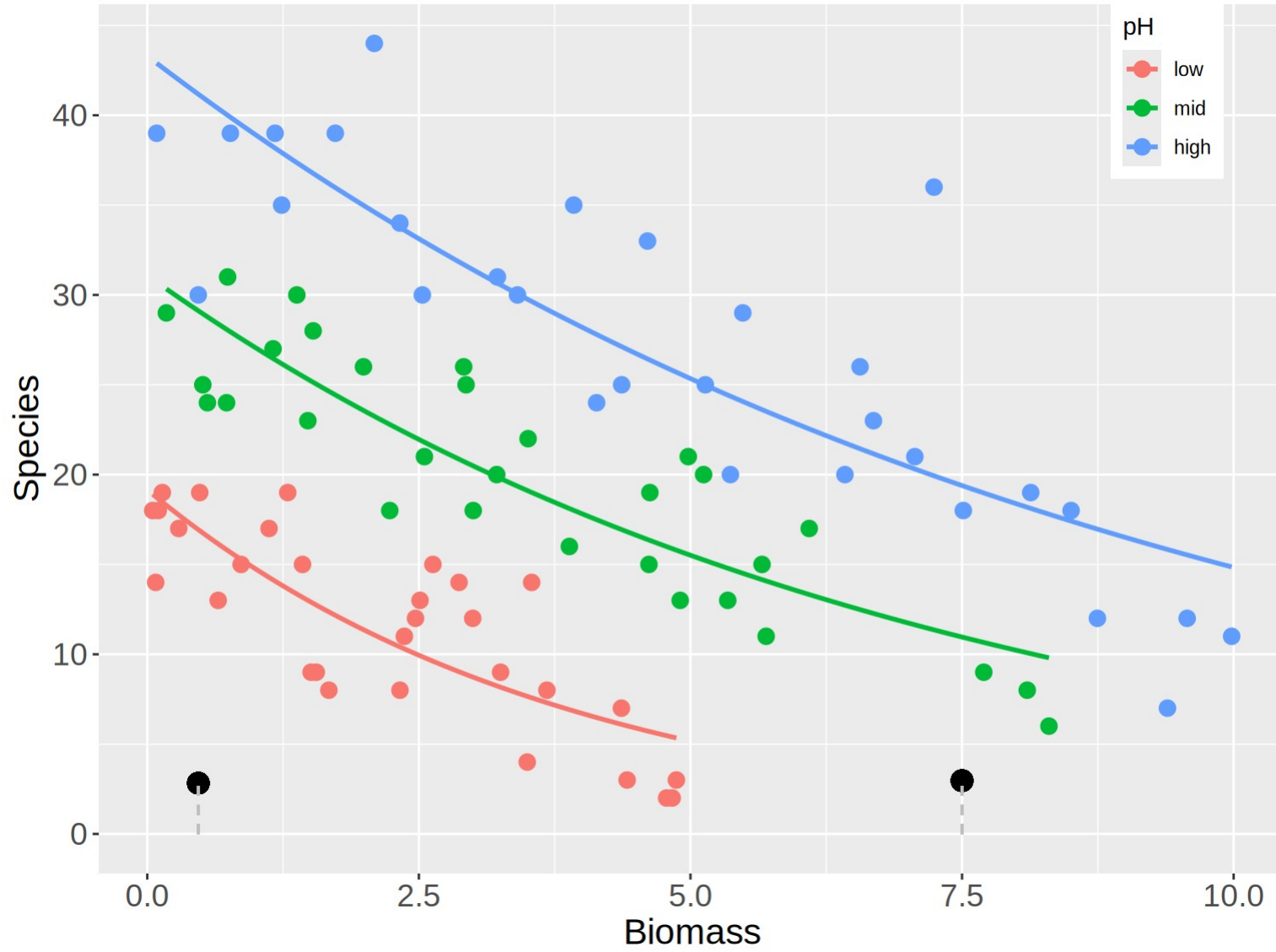
- **pH alto, biomassa 7.5**

```
## (Intercept)      Biomass      pHhigh Biomass:pHhigh
##      2.9525524    -0.2621580    0.8155712    0.1550282
```

$$2.95 - (0.26 * 7.5) + 0.82 + (0.15 * 7.5)$$

```
## [1] 2.96465
```

Predito pelo modelo



Preditor linear

$$\log(n_{sp}) = 2.95 - (0.26 * 7.5) + ($$

- baixo pH e biomassa 0.47: **2.83**
- alto pH e biomassa 7.50: **2.96**
- antilog: exp()

$$\exp(2.83)$$

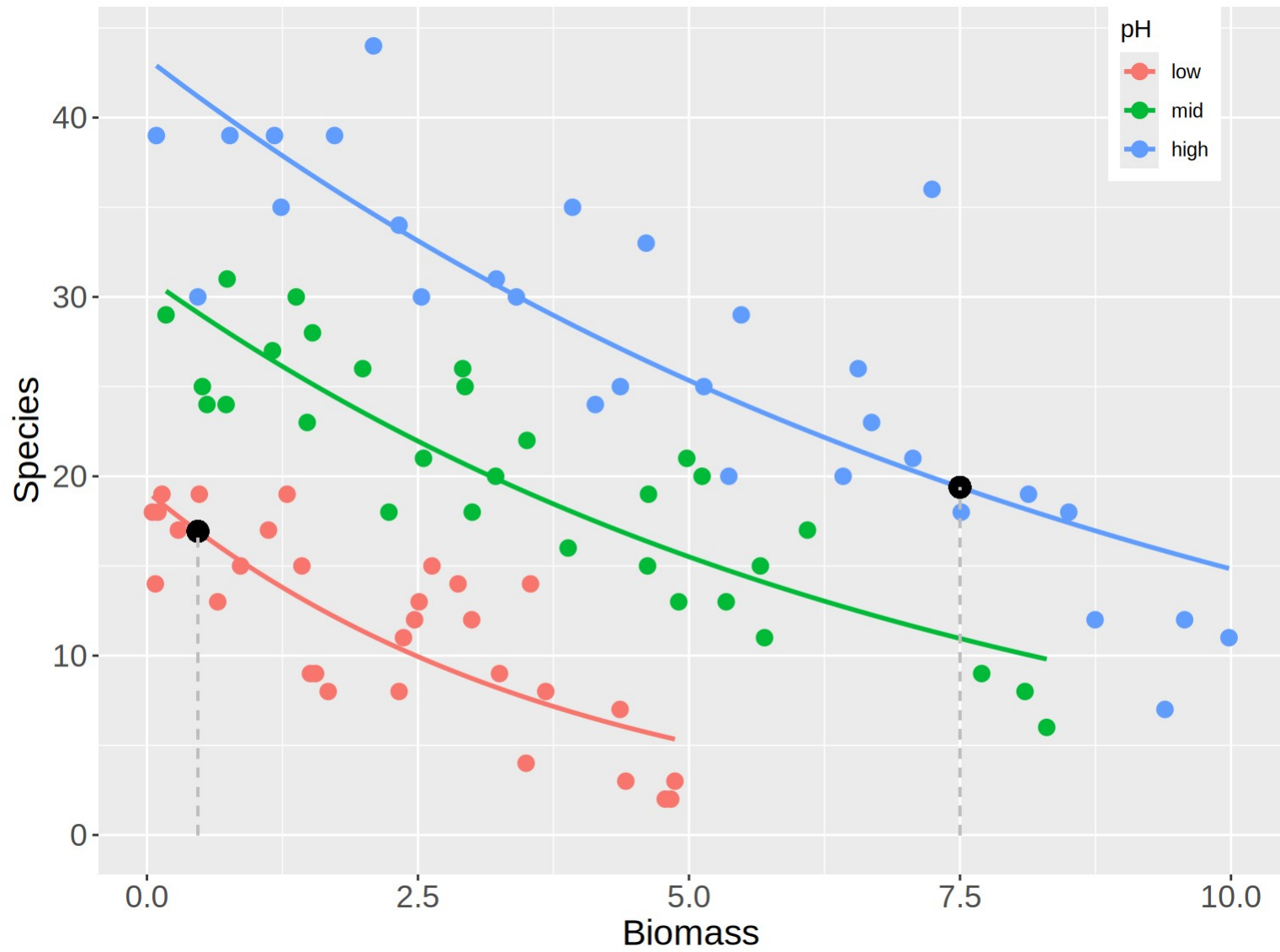
```
## [1] 16.93425
```

$$\exp(2.96)$$

```
## [1] 19.38792
```

Predito: função inversa

$$\exp(\eta) = \exp(\hat{\alpha} + \hat{\beta} * x)$$



summary(glm01)

```
#(Dispersion parameter for poisson family taken to be 1)  
# Null deviance: 452.346 on 89 degrees of freedom  
# Residual deviance: 83.201 on 84 degrees of freedom  
# AIC: 514.39
```

Fator dispersão

$$f_d = \frac{resid_{dev}}{df}$$

Sobredispersão

$$f_d \neq 1$$

- usar **family= quasipoisson**
- usar modelo com distribuição binomial negativa: **glm.nb()** pacote **MASS**

Interpretação dos coeficientes

```
coefglm01 <- coef(glm01)  
exp(coefglm01)
```

```
##      (Intercept)      Biomass      pHmid      pHhigh Biomass:pHmid  
##      19.1547812      0.7693895      1.6227280      2.2604666      1.1310384  
## Biomass:pHhigh  
##      1.1676909
```

GLM Contagem

- modelo cheio usando a família de ligação **poisson(log)**
- avalie a sobre-dispersão dos resíduos pela razão: **Residual deviance/degrees of freedom**
- se o valor da razão for muito maior que 1, ajuste o modelo com a família **quasipoisson**
- se a sobredispersão persistir: modelar o resíduo com a **binomial negativa** (um parâmetro a mais relacionado à agregação)
- compare os modelos simplificados com o mais complexo usando **anova**
 - com **poisson** use o argumento **test = "Chisq"**
 - com **quasipoisson** use o argumento **test = "F"**
- retenha o modelo mínimo adequado
- retorne os coeficientes e preditos do modelo para escala original (antilog)

Atividade

FIM