

# ME720 - Modelos Lineares Generalizados

## Parte 9 - Inferência em MLGs

Profa. **Larissa Avila Matos**

## Estimação dos parâmetros

# Estimação dos parâmetros

Uma vez definido cada componente do modelo, obteremos expressões gerais para a função de verossimilhança e para as distribuições assintóticas dos estimadores de máxima verossimilhança dos parâmetros para os MLGs.

Para  $n$  observações independentes, temos que a função log-verossimilhança do modelo é dada por  $\ell(\boldsymbol{\beta}, \phi) = \sum_{i=1}^n \ell_i(\boldsymbol{\beta}, \phi)$ , onde

$$\ell_i(\boldsymbol{\beta}, \phi) = \frac{[y_i \theta_i - b(\theta_i)]}{a(\phi)} + c(y_i, \phi).$$

Então,

$$\ell(\boldsymbol{\beta}, \phi) = \sum_{i=1}^n \frac{[y_i \theta_i - b(\theta_i)]}{a(\phi)} + \sum_{i=1}^n c(y_i, \phi).$$

Para um GLM  $\eta_i = \sum_{j=1}^p \beta_j x_{ij} = g(\mu_i)$ , com função de ligação  $g$ . Portanto, o sistema de equações de verossimilhança para  $\beta$  é dado por

$$\frac{\partial \ell(\beta, \phi)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \ell_i}{\partial \beta_j} = 0, \quad \forall j.$$

Para diferenciar a log-verossimilhança, usamos a regra da cadeia,

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{\partial \ell_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}.$$

Uma vez que,

$$\frac{\partial \ell_i}{\partial \theta_i} = \frac{[y_i - b'(\theta_i)]}{a(\phi)} \quad \text{e} \quad \mu_i = b'(\theta_i), \quad \text{Var}(Y_i) = b''(\theta_i)a(\phi),$$

temos que

$$\frac{\partial \ell_i}{\partial \theta_i} = \frac{(y_i - \mu_i)}{a(\phi)} \quad \text{e} \quad \frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i) = \frac{\text{Var}(Y_i)}{a(\phi)}.$$

Também uma vez que  $\eta_i = \sum_{j=1}^p \beta_j x_{ij}$ , então  $\frac{\partial \eta_i}{\partial \beta_j} = x_{ij}$ .

Finalmente,  $\eta_i = g(\mu_i) \Rightarrow \frac{\partial \mu_i}{\partial \eta_i}$ , depende da função de ligação para o modelo escolhido.

Resumindo, temos que

$$\begin{aligned} \frac{\partial \ell_i}{\partial \beta_j} &= \frac{\partial \ell_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \\ &= \frac{(y_i - \mu_i)}{a(\phi)} \frac{a(\phi)}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} = \frac{(y_i - \mu_i) x_{ij}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} \end{aligned}$$

A soma das  $n$  observações produz o sistema de equações de verossimilhança para um MLG.

# Equações de verossimilhança para um MLG

**Equações de verossimilhança para um MLG:**

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \beta_j} = \sum_{i=1}^n \frac{(y_i - \mu_i) x_{ij}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0, \quad j = 1, 2, \dots, p, \quad \text{e}$$

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \phi} = 0,$$

onde  $\eta_i = \sum_{j=1}^p \beta_j x_{ij} = g(\mu_i)$  para uma função de ligação  $g$ .

Seja  $V$  a matriz diagonal de variâncias das  $n$  observações, e seja  $D$  uma matriz diagonal com os elementos de  $\frac{\partial \mu_i}{\partial \eta_i}$ .

Para as expressões de um MLG  $\eta = X\beta$ , com matriz de planejamento  $X$ , as equações de verossimilhança têm a forma

$$XDV^{-1}(y - \mu) = 0.$$

Apesar de  $\beta$  não aparecer nessas equações, ele aparece implicitamente através de  $\mu_i = g^{-1}(\sum_{j=1}^p \beta_j x_{ij})$ .

Diferentes funções de ligação gera diferentes conjuntos de equações.

Essas equações são funções não lineares de  $\beta$ , e esse problema deve ser resolvido iterativamente. Há várias opções de algoritmos numéricos para resolução de sistemas não lineares: Newton-Raphson, Escore de Fisher (EF), Nelder-Mead, BFGS entre outros.

# Exercício

- Encontrar as equações de máxima verossimilhança para o Modelo Poisson Log Linear.



# Relação entre média e variância

Os sistema de equações de máxima verossimilhança depende da distribuição de  $Y_i$  somente pela média ( $\mathbb{E}(Y_i)$ ) e pela variância ( $\text{Var}(Y_i)$ ).

Além disso, a variância depende da média pela forma

$$\text{Var}(Y_i) = V(\mu_i),$$

para alguma função  $V(\cdot)$ .

Ou seja, a relação entre a média e a variância caracteriza a distribuição de  $Y_i$ .

**Exemplo:** Se  $Y_i$  tem distribuição pertencente a família exponencial e  $\text{Var}(Y_i) = \mu_i$ , então necessariamente  $Y_i$  tem distribuição de Poisson.

# Distribuição Assintótica de $\hat{\beta}$

Através das propriedades de máxima verossimilhança, e sob condições de regularidade, para  $n$  grande o estimador de máxima verossimilhança  $\hat{\beta}$  de  $\beta$  para um MLG é eficiente e tem distribuição Normal.

Temos que a matrix de informação  $\mathbf{I}$ , tem elementos dados por

$$\begin{aligned}\mathbb{E} \left( -\frac{\partial^2 \ell_i}{\partial \beta_j \partial \beta_k} \right) &= \mathbb{E} \left( \frac{\partial \ell_i}{\partial \beta_j} \frac{\partial \ell_i}{\partial \beta_k} \right) \\&= \mathbb{E} \left( \frac{(Y_i - \mu_i) x_{ij}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} \frac{(Y_i - \mu_i) x_{ik}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} \right) \\&= \frac{x_{ij} x_{ik}}{\text{Var}(Y_i)^2} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \underbrace{\mathbb{E}((Y_i - \mu_i)^2)}_{=\text{Var}(Y_i)} = \frac{x_{ij} x_{ik}}{\text{Var}(Y_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2.\end{aligned}$$

A matriz  $\mathbf{I}$  é chamada de matriz de informação esperada.

Então,

$$\mathbb{E} \left( -\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k} \right) = \sum_{i=1}^n \frac{x_{ij} x_{ik}}{\text{Var}(Y_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

Seja,  $\mathbf{W}$  uma matriz diagonal com elementos dados por

$$w_i = \frac{(\partial \mu_i / \partial \eta_i)^2}{\text{Var}(Y_i)}$$

$$\Rightarrow \mathbf{I} = \mathbf{X}' \mathbf{W} \mathbf{X}$$

A forma de  $\mathbf{W}$  depende da função de ligação  $g(\cdot)$ , uma vez que  $\partial \mu_i / \partial \eta_i = g'(\mu_i)$ .

Portanto, a matriz de covariância de  $\hat{\boldsymbol{\beta}}$  é dada pela inversa da matriz de informação  $\mathbf{I}$ .

**Distribuição Assintótica de  $\hat{\beta}$  para um MLG  $\eta = X\beta$ :**

$\hat{\beta}$  tem distribuição aproximadamente normal,  $N_p(\beta, (X'WX)^{-1})$ ,

onde  $W$  é a matriz diagonal com elementos  $w_i = \frac{(\partial\mu_i/\partial\eta_i)^2}{\text{Var}(Y_i)}$ .

A matriz de covariância assintótica é estimada por  $\widehat{\text{Var}}(\hat{\beta}) = (X'\widehat{W}X)^{-1}$ , onde  $\widehat{W}$  é  $W$  avaliado em  $\hat{\beta}$ .

**Obs:**

- 1 Para a FE com parâmetro de escala,  $\theta$  e  $\phi$  são parâmetros ortogonais.
- 2  $\hat{\beta}$  e  $\hat{\phi}$  são assintoticamente independentes.

# Matriz de covariância para os valores ajustados

O preditor linear estimado é dado por

$$\hat{\eta} = \mathbf{X}\hat{\beta}.$$

Para  $n$  grande, temos

$$\text{Var}(\hat{\eta}) = \mathbf{X}'\text{Var}(\hat{\beta})\mathbf{X} \approx \mathbf{X}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}.$$

Podemos obter a variância assintótica de  $\hat{\mu}$  ( $\text{Var}(\hat{\mu})$ ) por  $\text{Var}(\hat{\eta})$ , através do método delta. Então,

$$\text{Var}(\hat{\mu}) \approx \mathbf{D}'\text{Var}(\hat{\eta})\mathbf{D} \approx \mathbf{D}\mathbf{X}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}\mathbf{D},$$

onde  $\mathbf{D}$  é uma matriz diagonal com elementos  $\partial\mu_i/\partial\eta_i$ .

**Exercício:** Pesquisar sobre o método delta.

## Exercício

- Voltando ao exemplo da Poisson, encontre os elementos da matriz  $\mathbf{W}$ .

# Estimação de $\beta$

Como encontramos os estimadores de máxima verossimilhança dos parâmetros de um MLG?

**Problema:** O sistema de equações são geralmente não-lineares em  $\beta$ .

**Solução:** Métodos iterativos para resolver sistema de equações não lineares. Focaremos em dois métodos:

- *Newton-Raphson*

- *Escore de Fisher*

# Método de Newton-Raphson

O algoritmo Newton-Raphson, método de Newton-Raphson, foi desenvolvido por Isaac Newton e Joseph Raphson e tem o objetivo estimar as raízes de uma função.

Suponha que queremos encontrar a solução da equação  $g(x_0) = 0$ , onde  $g$  é uma função diferenciável. Dado um número  $x$  próximo de  $x_0$ , segue da expansão em série de Taylor em torno de  $x$  que

$$0 = g(x_0) \approx g(x) + g'(x)(x_0 - x).$$

Resolvendo para  $x_0$ , temos

$$x_0 \approx x - \frac{g(x)}{g'(x)}.$$

Assim, dado um valor estimado  $x_t$ , então podemos ter um novo valor estimado  $x_{t+1}$  por

$$x_{t+1} \approx x_t - \frac{g(x_t)}{g'(x_t)}.$$

Este procedimento é repetido para  $t = 1, 2, 3, \dots$  até  $|g(x_t)/g'(x_t)|$  ser suficientemente pequeno.



# Método de Newton-Raphson

Voltando ao nosso problema, queremos encontrar a solução da equação  $\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}$ .

No passo  $t$ , o processo iterativo ( $t = 0, 1, 2, \dots$ ) aproxima  $\ell(\boldsymbol{\beta})$  próximo de  $\boldsymbol{\beta}^{(t)}$  pela expansão em série de Taylor de segunda ordem,

$$\ell(\boldsymbol{\beta}) \approx \ell(\boldsymbol{\beta}^{(t)}) + \mathbf{u}^{(t)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}) + \left(\frac{1}{2}\right) (\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)})' \mathbf{H}^{(t)} (\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}),$$

onde  $\mathbf{u}^{(t)}$  e  $\mathbf{H}^{(t)}$  são  $\mathbf{u}$  e  $\mathbf{H}$  avaliados em  $\boldsymbol{\beta}^{(t)}$  respectivamente, com

■  $\mathbf{u} = \left( \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_1}, \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_2}, \dots, \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_p} \right)'$ , e

■  $\mathbf{H}$  a matriz *Hessiana*, onde  $H_{ij} = \frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j}$ .

**Obs:**  $\boldsymbol{\beta}^{(0)}$  valor inicial (chute inicial).

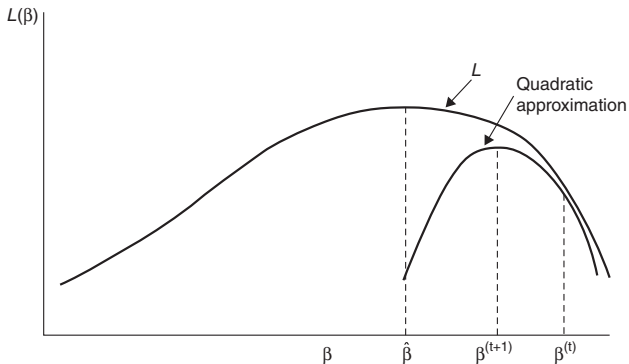
Resolvendo,  $\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \approx \mathbf{u}^{(t)} + \mathbf{H}^{(t)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}) = \mathbf{0}$ , temos a seguinte aproximação

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - (\mathbf{H}^{(t)})^{-1} \mathbf{u}^{(t)},$$

assumindo que  $\mathbf{H}^{(t)}$  é não singular.

O procedimento descrito é repetido até que mudanças em  $\ell(\boldsymbol{\beta}^{(t)})$  entre ciclos sucessivos são suficientemente pequenas.

Para muitos MLGs, a matriz Hessiana é negativa definida, e a log verossimilhança é uma função estritamente côncava. Então, as estimativas de máxima verossimilhança dos parâmetros do modelo existem e são únicas sob condições bastante gerais. A convergência de  $\boldsymbol{\beta}^{(t+1)}$  para  $\hat{\boldsymbol{\beta}}$  na vizinhança de  $\hat{\boldsymbol{\beta}}$  é rápida.



**Figure 4.2** Illustration of a cycle of the Newton–Raphson method.

Figura ilustrativa do livro texto (Agresti, A. (2015)).

# Método de Escore de Fisher

O método de Escore de Fisher é um método iterativo alternativo para resolver o sistema de equações de verossimilhança.

A diferença entre o método de Escore de Fisher e o método de Newton-Raphson está na maneira como escolhemos a matriz *Hessiana*.

O método de Escore de Fisher usa o valor esperado da matriz *Hessiana*, chamada de matriz de informação esperada, enquanto método de Newton-Raphson usa a própria matriz *Hessiana*, chamada de matriz de informação observada.

Portanto, temos a seguinte aproximação

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + (\mathbf{I}^{(t)})^{-1} \mathbf{u}^{(t)},$$

onde  $\mathbf{I}^{(t)}$  é  $\mathbf{I}$  avaliado em  $\boldsymbol{\beta}^{(t)}$ , ou seja  $\mathbf{I}^{(t)}$  tem elementos  $-\mathbb{E}(\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j})$  avaliado em  $\boldsymbol{\beta}^{(t)}$ .

Os algoritmos de Escore de Fisher e de Newton-Raphson são idênticos para os MLGs que usam a função de ligação canônica (Nelder e Wedderburn, 1972).

# Newton-Raphson e Escore de Fisher

Newton-Raphson e Escore de Fisher para o parâmetro de uma binomial

Exemplo para o método de Newton-Raphson e Escore de Fisher de um problema mais simples para o qual sabemos a resposta.

# Newton-Raphson e Escore de Fisher

Newton-Raphson e Escore de Fisher para o parâmetro de uma binomial

Exemplo para o método de Newton-Raphson e Escore de Fisher de um problema mais simples para o qual sabemos a resposta.

Seja  $Y$  uma a.a de uma distribuição  $Bin(n, p)$ , a função de log-verossimilhança é dada por

$$\ell(p) = \log(p^{nY}(1-p)^{n-nY}) = ny \log(p) + (n - ny) \log(1-p).$$

Sabemos que

$$u = \frac{\partial \ell}{\partial p} = \frac{(ny - np)}{p(1-p)} \quad \text{e} \quad H = \frac{\partial^2 \ell}{\partial p^2} = - \left[ \frac{ny}{p^2} + \frac{n - ny}{(1-p)^2} \right].$$

Maximizando a log-verossimilhança temos que o estimador de MV para  $p$  é  $\hat{p} = y$ .

Cada passo do algoritmo de Newton-Raphson é dado por

$$p^{(t+1)} = p^{(t)} + \left[ \frac{ny}{(p^{(t)})^2} + \frac{n - ny}{(1 - p^{(t)})^2} \right]^{-1} \frac{(ny - np^{(t)})}{p^{(t)}(1 - p^{(t)})}.$$

- Se,  $p^{(0)} = \frac{1}{2} \Rightarrow p^{(1)} = y$ .
- Quando  $p^{(t)} = y$ , temos que  $p^{(t+1)} = y$ , ou seja, não temos nenhuma alteração da iteração  $t$  para  $t + 1$ , o qual é o estimador de MV.



Calculando a esperança de  $H$ , temos que

$$I = \frac{n}{p(1-p)},$$

e cada passo do algoritmo de Escore de Fisher é dado por

$$\begin{aligned} p^{(t+1)} &= p^{(t)} + \left[ \frac{n}{p^{(t)}(1-p^{(t)})^2} \right]^{-1} \frac{(ny - np^{(t)})}{p^{(t)}(1-p^{(t)})} \\ &= p^{(t)} + (y - p^{(t)}) = y \end{aligned}$$

- Ou seja,  $p^{(t+1)}$  é o estimador de MV após apenas uma iteração e ficando nesse valor em todas as iterações.

# Estimação do parâmetro de escala

- 1 Maximizar  $\ell(\beta, \phi)$  com respeito a  $\phi$ . (Muito sensível a suposição da distribuição)
- 2 Sabemos que  $\text{Var}(Y_i) = a(\phi)V(\mu_i)$ , então

$$\frac{\mathbb{E}((Y_i - \mu_i)^2)}{V(\mu_i)} = a(\phi)$$

$$\Rightarrow \hat{a}(\phi) = \frac{1}{n-p} \sum_{i=1}^n \frac{(Y_i - \hat{\mu}_i)^2}{\text{Var}(\hat{\mu}_i)}.$$

Vimos que  $\hat{\beta}$  tem distribuição aproximadamente normal, i.e.,

$$\hat{\beta} \approx N_p(\beta, (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}),$$

onde  $\mathbf{W}$  é a matriz diagonal com elementos  $w_i = \frac{(\partial\mu_i/\partial\eta_i)^2}{\text{Var}(Y_i)}$ , com  $\text{Var}(Y_i) = b''(\theta_i)a(\phi)$ .

Se  $\hat{\beta}_j$  é a  $j$ -ésima componente do vetor  $\hat{\beta}$ , então

$$\hat{\beta}_j \approx N_p(\beta_j, \psi_j),$$

em que  $\psi_j$  é o  $j$ -ésimo elemento da diagonal principal de  $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$ .

Vimos também que  $\widehat{\beta}$  é um estimador consistente, então

$$\frac{\widehat{\beta}_j - \beta_j}{\sqrt{\widehat{\psi}_j}} \approx N(0, 1),$$

onde  $\widehat{\psi}_j$  é o j-ésimo elemento da diagonal principal de  $(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1}$ . Veremos mais adiante que  $(\widehat{\beta}_j - \beta_j)/\sqrt{\widehat{\psi}_j}$  é a estatística de Wald.

Portanto, um intervalo  $(1 - \alpha)\%$  de confiança para  $\beta_j$  é dado por

$$IC(100(1 - \alpha)\%, \beta_j) = \left[ \widehat{\beta}_j - z_{\frac{\alpha}{2}} \sqrt{\widehat{\psi}_j}; \quad \widehat{\beta}_j + z_{\frac{\alpha}{2}} \sqrt{\widehat{\psi}_j} \right].$$

## Exemplo

# Exemplo: Exposição de Bactérias

Vamos considerar um exemplo de um modelo log-linear de Poisson para ajustar dados de contagem.

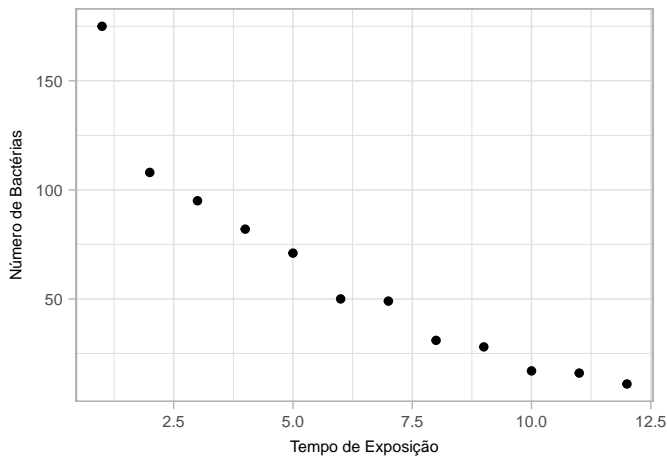
Recordando - Variáveis:

- variável resposta: número de bactérias sobreviventes em amostras de um produto alimentício exposto a uma temperatura de 300°F.
- variável explicativa: tempo de exposição do produto (em minutos).

(Montgomery, Peck e Vining, 2001) (Paula, 2013a).

## Descrição dos Dados:

Bactérias	175	108	95	82	71	50	49	31	28	17	16	11
Exposição	1	2	3	4	5	6	7	8	9	10	11	12



# Exemplo: Exposição de Bactérias

Ajuste Modelo Log-linear de Poisson:

$$\log(\mu_i) = \alpha + \beta \text{ tempo}_i,$$

em que  $y_i \sim P(\mu_i)$ . As estimativas desse modelo usando a função `glm()` do R é dada a seguir.

```
fit<-glm(bacterias~exposicao,family=poisson(link = "log"))  
summary(fit)$coefficients
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	5.3055715	0.06348273	83.57504	0.000000e+00
exposicao	-0.2288956	0.01269994	-18.02336	1.277325e-72



Iremos agora implementar a “nossa” função `glm` do modelo log-linear de Poisson através do método de Newton- Raphson. Encontraremos:

- as estimativas de  $\beta$ ,
- a matriz de covariância estimada de  $\beta$ ,
- testes de hipóteses simples, e
- intervalos de confiança.

**Obs.:** Para o modelo log-linear de Poisson o método de Newton- Raphson é equivalente ao método de Ecore de Fisher.

```

newton <-function(y,X,init,eps=1e-6,maxiter=50){
  beta <- init
  n<-dim(X)[1]
  out <- matrix(NA, nrow=maxiter+1,ncol=length(t(beta)))
  out[1,] <- t(init)
  i <- 1
  continue <- T
  while (continue) {
    i <- i+1
    beta.o <- beta
    W<-diag(n)
    diag(W)<-exp(X%%beta.o)
    mu<-exp(X%%beta.o)
    beta <- beta.o + solve(t(X)%*%W%*%X)%*%t(X)%*%(y-mu)
    if(sum(is.na(beta))>0){stop("NA nas estimativas")}
    out[i,] <- t(beta)
    continue <- (abs(beta-beta.o) > eps) && (i <= maxiter)
  }
  if (i > maxiter) {
    warning("Máximo número de iterações atingido")
  }
  out <- out[!is.na(out[,1]),]
  saida<-list(out=out,est=out[i,],iter=i)
  return(saida)
}

```

```
chute.inicial<-matrix(c(1,1),2,1)
intercepto<-rep(1,length(exposicao))
X<-cbind(intercepto,exposicao)
newton.posssoin<-newton(bacterias,X,chute.inicial,eps=1e-6,maxiter=20)
newton.posssoin
```

\$out

	[,1]	[,2]
[1,]	1.00000000	1.00000000
[2,]	0.09664021	0.99162795
[3,]	-0.64604637	0.96932626
[4,]	-0.98475052	0.91193810
[5,]	-0.43927926	0.77734889
[6,]	1.37981370	0.52942633
[7,]	3.33639853	0.26268220
[8,]	4.26155271	0.08280845
[9,]	4.82694578	-0.06209716
[10,]	5.14919441	-0.16720809
[11,]	5.28158348	-0.21856597
[12,]	5.30484337	-0.22856887
[13,]	5.30557072	-0.22889523
[14,]	5.30557147	-0.22889557

\$est

[1] 5.3055715 -0.2288956

\$iter

[1] 14

```
estimativas<-newton.poisson$est  
estimativas
```

```
[1] 5.3055715 -0.2288956
```

```
W<-diag(length(exposicao))  
diag(W)<-exp(X%*%estimativas)  
I=t(X)%*%W%*%X  
se=sqrt(diag(solve(I)))  
se
```

```
intercepto  exposicao  
0.06348273  0.01269994
```

```
z.value=estimativas/se  
z.value
```

```
intercepto    exposicao  
83.57504      -18.02336
```

```
p.value=pnorm(abs(z.value), lower.tail = F)  
p.value
```

```
intercepto    exposicao  
0.000000e+00  6.386627e-73
```

```
cbind(estimativas,se,z.value,p.value)
```

	estimativas	se	z.value	p.value
intercepto	5.3055715	0.06348273	83.57504	0.000000e+00
exposicao	-0.2288956	0.01269994	-18.02336	6.386627e-73

```
summary(fit)$coefficients
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	5.3055715	0.06348273	83.57504	0.000000e+00
exposicao	-0.2288956	0.01269994	-18.02336	1.277325e-72

```
z.alpha<-qnorm(0.975)
LI=estimativas-z.alpha*se
LI
```

```
intercepto    exposicao
    5.181148   -0.253787
```

```
LS=estimativas+z.alpha*se
LS
```

```
intercepto    exposicao
    5.4299953  -0.2040042
```

$$IC(95\%, \beta_0) = [5.1811476; 5.4299953]$$

$$IC(95\%, \beta_1) = [-0.253787; -0.2040042]$$

Usando o estimador de minimos quadrados do modelo linear normal como valor inicial.

```
EMQ<-matrix(solve(t(X)%*%X)%*%t(X)%*%bacterias,2,1)  
EMQ
```

```
      [,1]  
[1,] 142.19697  
[2,] -12.47902
```

```
n1<-newton(bacterias,X,EMQ,eps=1e-6,maxiter=20)
```

Warning in newton(bacterias, X, EMQ, eps = 1e-06, maxiter = 20): Máximo número de iterações atingido

```
n1$est
```

```
[1] 122.19697 -12.47902
```



	[,1]	[,2]
[1,]	142.197	-12.47902
[2,]	141.197	-12.47902
[3,]	140.197	-12.47902
[4,]	139.197	-12.47902
[5,]	138.197	-12.47902
[6,]	137.197	-12.47902
[7,]	136.197	-12.47902
[8,]	135.197	-12.47902
[9,]	134.197	-12.47902
[10,]	133.197	-12.47902
[11,]	132.197	-12.47902
[12,]	131.197	-12.47902
[13,]	130.197	-12.47902
[14,]	129.197	-12.47902
[15,]	128.197	-12.47902
[16,]	127.197	-12.47902
[17,]	126.197	-12.47902
[18,]	125.197	-12.47902
[19,]	124.197	-12.47902
[20,]	123.197	-12.47902
[21,]	122.197	-12.47902

```
n2<-newton(bacterias,X,EMQ,eps=1e-6,maxiter=200)
n2$iter
```

```
[1] 130
```

```
n2$est
```

```
[1] 5.3055715 -0.2288956
```

Valores iniciais com problemas:

```
newton(bacterias,X,matrix(c(-1,-1),2,1),eps=1e-6,maxiter=20)
```

```
Error in newton(bacterias, X, matrix(c(-1, -1), 2, 1), eps = 1e-06, maxiter = 20): NA nas estimativas
```

```
EMQ1<-matrix(solve(t(X)%*%X)%*%t(X)%*%sqrt(bacterias),2,1)
newton(bacterias,X,EMQ1,eps=1e-6,maxiter=20)
```

\$out

	[,1]	[,2]
[1,]	12.572735	-0.8160687
[2,]	11.566699	-0.8109121
[3,]	10.550649	-0.7971663
[4,]	9.509567	-0.7617262
[5,]	8.414967	-0.6781443
[6,]	7.255059	-0.5215246
[7,]	6.185418	-0.3472236
[8,]	5.528580	-0.2561283
[9,]	5.322676	-0.2309032
[10,]	5.305679	-0.2289080
[11,]	5.305571	-0.2288956
[12,]	5.305571	-0.2288956

\$est

[1] 5.3055715 -0.2288956

\$iter

[1] 12

- [Notas](#) de aula do Prof. Gilberto de Paula.
- Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Wiley series in probability and statistics.