

Lista 5 - MAE0524

Guilherme Navarro N^oUSP:8943160 Leonardo Noronha N^oUSP:9793436

Exercício 3

Deseja-se ajustar o modelo não linear $\beta_0 x^{\beta_1}$, relacionando o peso (y) com o comprimento (x) de determinada espécie de peixe. Para isso será utilizada uma linearização do modelo dada por:

$$\log(y) = \beta_0^* + \beta_1 \log(x) + e$$

Em que $\beta_0^* = \log(\beta_0)$ e $e \sim N(0, \sigma^2)$

Na tabela abaixo apresentamos o comprimento (em cm) e o peso (em kg) de 25 exemplares de um peixe capturados na costa sul do Brasil.

x	10,2	19,0	22,3	23,8	24,9	26,1	26,9	27,9	28,5	29,2	30,1	31,0	31,7
y	0,01	0,09	0,12	0,18	0,19	0,23	0,24	0,31	0,33	0,32	0,33	0,40	0,45
x	32,4	32,9	33,3	34,1	34,6	35,5	36,3	37,0	38,0	39,0	39,6	43,0	
y	0,49	0,67	0,48	0,54	0,63	0,59	0,65	0,50	0,81	0,87	0,71	0,91	

- (a) Construa os diagramas de dispersão de y versus x e de $\log(y)$ versus $\log(x)$. Comente

Resolução

Gráfico 1: Diagrama de dispersão

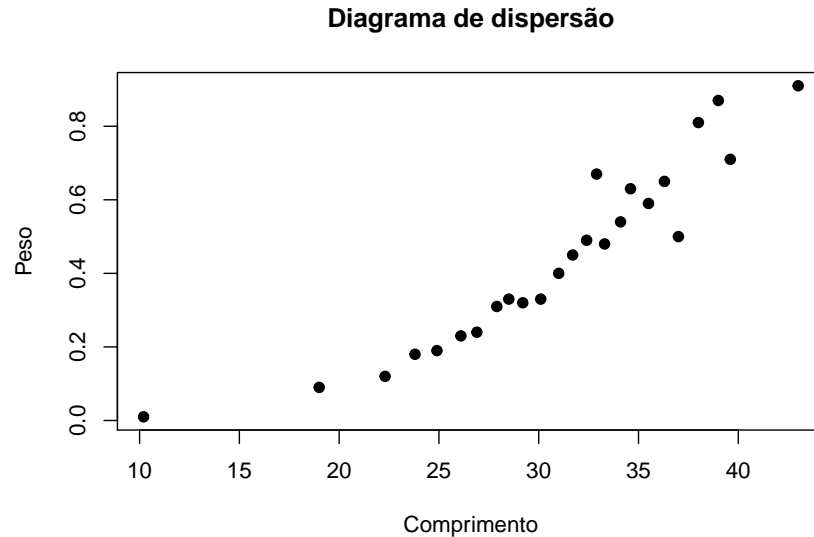
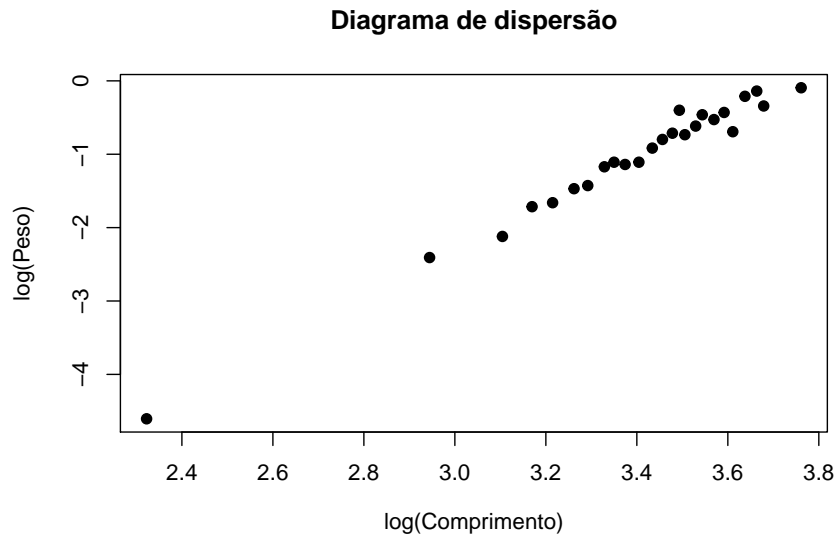


Gráfico 2: Diagrama de dispersão com as variáveis com logaritmo aplicado



Como podemos observar no gráfico 1 temos uma tendência não-linear, e quando fizemos o logaritmo nas variáveis comprimento e peso no gráfico 2 corrigimos a falta de linearidade, como vantagem nós podemos ajustar um modelo linear, como desvantagem as variáveis ficam com a unidade alterada pelo logaritmo, ou seja, pode ter uma interpretação difícil.

- (b) Considerando *a priori* de Jeffreys, especifique a distribuição a posteriori conjunta de (β_0^*, β_1) . Simule uma amostra de tamanho $M = 1000$ desta distribuição.

Resolução

Considerando *a priori* de Jeffreys, temos:

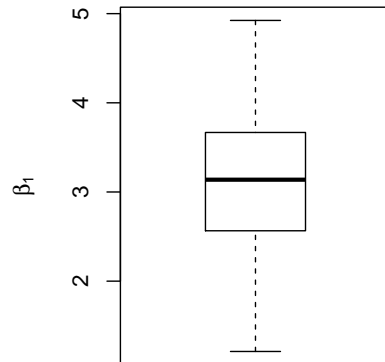
```
library(LearnBayes)

logpost <- function(theta,data){ # log da posteriori de theta
  beta.0 <- theta[1]
  beta.1 <- theta[2]
  x <- data[,1]
  y <- data[,2]
  n <- length(x)
  -1/2*sum((y-beta.0-beta.1*x)^2)
}

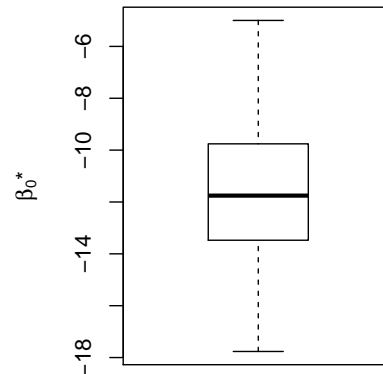
laplace <- laplace(logpost,c(-12,3),data) # aprox pela normal

proposal <- list(var=laplace$var, scale=2)
start <- array(c(-12,3),c(1,2)) # chute inicial
m <- 1000 # tamanho da amostra
s <- rwmetro(logpost, proposal, start,m,data)
```

Box-Plot dos valores simulado:



Box-Plot dos valores simulado:



- (c) Com base na amostra simulada em (b), obtenha uma amostra de (β_0, β_1) . Aproxime as distribuições a posteriori marginais via histograma dos dados simulados. Obtenha os intervalos de credibilidade 0.90 para β_0 e β_1 .

Resolução

Com base na amostra simulado no item b, temos que $\beta_0^* = \log(\beta_0)$ basta fazer a seguinte transformação nos dados $\beta_0 = e^{\beta_0^*}$, assim temos que o intervalo de credibilidade de 90% para β_1 é:

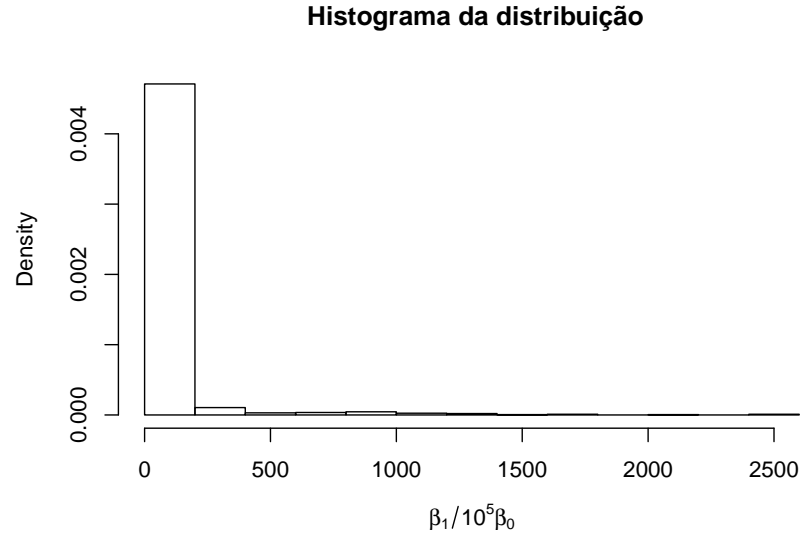
```
##           5%           95%
## 1.882094 4.301029
```

E o intervalo de credibilidade para β_0 é:

```
##           5%           95%
## 1.430230e-07 5.290431e-04
```

- (d) Considere que a razão $\frac{\beta_1}{10^5 \beta_0}$ seja de interesse para estudos sobre biologia da espécie. Aproxime a distribuição a posteriori dessa quantidade via histograma, obtenha estimativas pontuais para ela, bem como, o intervalo de credibilidade 0.90.

Resolução



```
##           5%           95%
## 0.03560268 301.66655138
```

Exercício 4

Considere um experimento onde os indivíduos indicam o número de eventos estressantes (y_i) aos quais foram submetidos em determinado mês (i). Os dados são apresentados na tabela a seguir

Mês	1	2	3	4	4	5	6	7	8	9	10	11	12	13	14	15	17	18
y	15	11	14	17	5	11	10	4	8	10	7	9	11	3	6	1	1	4

Vamos considerar o modelo de regressão loglinear Poisson:

$$\log(\lambda_i) = \beta_0 + \beta_1 i$$

em que $Y_i | \lambda_i \sim \text{Poisson}(\lambda_i)$

- (a) Sob distribuição a priori uniforme (imprópria) para (β_0, β_1) . Escreva uma função no **R** para calcular o logaritmo da posteriori.

Resolução

Sendo a função log-verossimilhança para esse modelo:

$$\log(L(\lambda, Y)) = \sum_{i=1}^{18} [y_i(\beta_0 + \beta_1 i) - e^{(\beta_0 + \beta_1 i)}]$$

```

data.x <- c(seq(1:18))
data.y <- c(15,11,14,17,5,11,10,4,8,10,7,9,11,3,6,1,1,4)
Data <- data.frame(data.x,data.y)

log.post <- function(theta,data){ # log da posteriori de theta
  beta.0 <- theta[1]
  beta.1 <- theta[2]
  i <- data[,1]
  y <- data[,2]
  sum(y*(beta.0 + beta.1*i) - exp(beta.0 + beta.1*i))
}

```

- (b) Use o algoritmo de metropolis com proposta passeio aleatório para simular uma amostra de tamanho 1000 da distribuição a posteriori de β_1 . Apresente os valores simulados usando o gráfico box-plot.

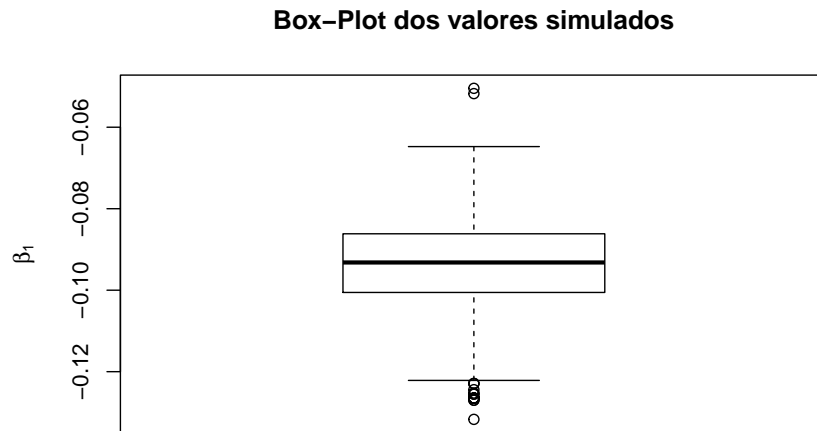
Resolução

```

library(LearnBayes)
laplace <- laplace(log.post,c(9.5,5),Data)
proposal <- list(var=laplace$var, scale=2)
start <- array(c(2.87,-0.03),c(1,2))
m <- 1000
s <- rwmetrop(log.post, proposal, start,m,Data)

boxplot(s$par[,2], ylab=expression(beta[1]), main="Box-Plot dos valores simulados")

```



(c) Obtenha um intervalo de credibilidade 0.9 para β_1 .

Resolução

```
quantile(s$par[,2],c(.05,.95))
```

```
##           5%           95%  
## -0.11440774 -0.07453924
```

(d) Considerando o intervalo construído em (c), você diria que há evidências para rejeitar a hipótese $H_0 : \beta_1 = 0$?

Resolução

Com base no item C, podemos notar que o *zero* não está contido no intervalo de credibilidade, assim temos evidências estatísticas para rejeitar $H_0 : \beta_1 = 0$ a um nível de significância de 10%.

Exercício 6

Seja $Y|p \sim \text{bin}(n, p)$ e $\theta = \log(\frac{p}{1-p})$. Considere θ com distribuição a priori Normal com média zero e desvio padrão 0.25. Para $n = 5$ e $y = 5$, determine a probabilidade a posteriori de $\theta > 0$ usando as duas aproximações indicadas a seguir. (Note que $\theta > 0$ equivale a $p > 0.5$)

(a) Use a aproximação normal para a densidade a posteriori

Resolução

```
log.post <- function(theta,data){ # log da posteriori de theta  
  y <- data  
  n <- 5  
  k <- length(data)  
  theta*(sum(y)-8*theta)-n*k*log(1+exp(theta))  
}
```

```
dados <- c(5)  
aproxnorm <- laplace(log.post,0.5,dados)  
aproxnorm
```

```
## $mode  
## [1] 0.1449219  
##  
## $var  
##           [,1]  
## [1,] 0.057993  
##  
## $int  
## [1] -3.789343
```

```
##
## $converge
## [1] TRUE
```

Probabilidade de $\theta > 0$:

```
## [1] 0.7263436
```

- (b) Use o algoritmo de Metropolis-Hasting com proposta passeio aleatório. No algoritmo considere o desvio padrão da proposta igual a duas vezes o desvio padrão da aproximação normal obtida em (a).

Resolução

Sendo a distribuição a posteriori: $\pi(\theta|Y) \propto L(Y, \theta) * f(\theta)$ Com

$$L(Y, \theta) \propto e^{\theta \sum_{i=1}^k y_i} (1 + e^{\theta})^{-nk}$$

com n é fixado igual a 5, e k o tamanho da amostra. E $f(\theta) = \frac{4}{\sqrt{2\pi}} e^{-8\theta^2}$. Assim a distribuição a posteriori proporcional é:

$$\pi(\theta|Y) \propto e^{\theta(\sum_{i=1}^k y_i - 8\theta)} (1 + e^{\theta})^{-nk}$$

```
start <- array(c(aproxnorm$mode, 2*sqrt(aproxnorm$var)), c(1,1))
proposal <- list(var=2*aproxnorm$var, scale=2)
m <- 10000 # tamanho da amostra
s <- rwmthrop(log.post, proposal, start, m, dados)
```

Probabilidade de $\theta > 0$:

```
## [1] 0.718
```