Trabalho3_Markdown

Rodrigo Malta Esteves

31/03/2021

1. Etapa 0: Análise preliminar dos dados
2. Etapa 1: Estimação de θ - obtenção da distribuição a posteriori e sua exploração.
3. Etapa 2: Exploração do comportamento de uma observação futura e obtenção de distribuição preditiva.

```
library(tidyverse)
```

Etapa 0: Análise preliminar dos dados

Observamos as dimensões, definimos os nomes das colunas para melhor entendimento e observamos os histogramas relacionados a cada uma das colunas.

```
dados <- read.table("C:\\Users\\malta\\Desktop\\Pós Graduação\\Introdução ao Aprendizado Estatístico\\T
names(dados) <- c("custos","log_custos")
head(dados) #Visualização das primeiras linhas do arquivo lido
```

```
##     custos log_custos
## 1 6.846861   1.923790
## 2 6.947157   1.938333
## 3 5.788298   1.755838
## 4 6.927557   1.935507
## 5 9.035326   2.201142
## 6 8.192294   2.103194
```

```
dim(dados) #Conferindo a dimensão do arquivo
```

```
## [1] 1000    2
```

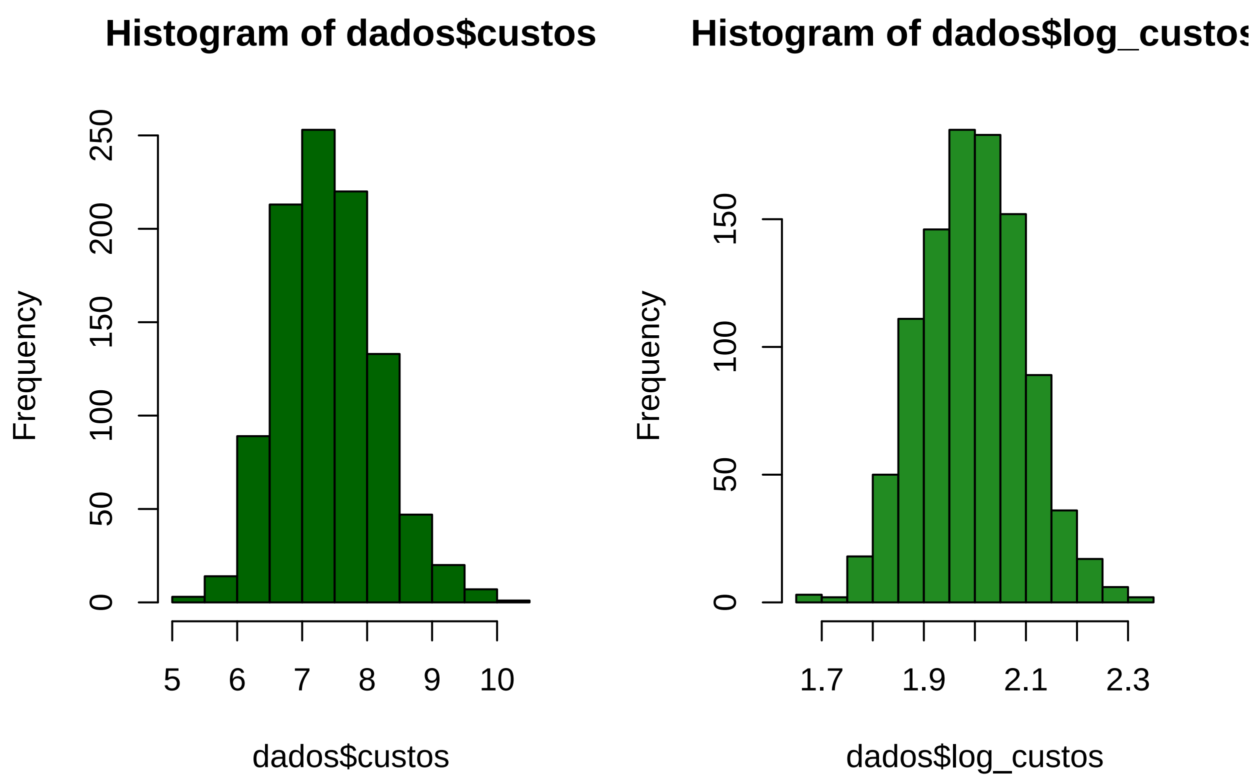
```
class(dados) #Tipo do arquivo
```

```
## [1] "data.frame"
```

```
summary(dados) #Resumo dos dados
```

```
##      custos         log_custos
##  Min.   : 5.249   Min.   :1.658
##  1st Qu.: 6.865   1st Qu.:1.926
##  Median : 7.342   Median :1.994
##  Mean   : 7.397   Mean   :1.996
##  3rd Qu.: 7.876   3rd Qu.:2.064
##  Max.   :10.100   Max.   :2.313
```

```r
par(mfrow=c(1,2))
hist(dados$custos,col="darkgreen")
hist(dados$log_custos,col="forestgreen")
```



Etapa 1: Estimação de θ - obtenção da distribuição a posteriori e sua exploração.

a) Distribuição da posteriori de theta

Observamos que a média da distribuição a posteriori será igual a média da verossimilhança, uma vez que a distribuição priori tem média igual a 0. Já o desvio padrão se tornará muito pequeno, dado que tamanho da amostra torna a precisão a posteriori extremamente grande.

```
# Medidas amostrais
t <- sum(dados$log_custos) #Somatório
phi<-1/(sd(dados$log_custos))^2 #Precisão
xbarra<-mean(dados$log_custos) #Média
n<-length(dados$log_custos) #Número de observações
par_vero <- cbind(xbarra,1/phi)

#Priori -> theta~Normal(0,10^2):
m.0<-0
d.0<-10
phi.0<-1/d.0^2
par_pri <- cbind(m.0,1/phi.0)

#Posteriori:
phi.n<-phi.0+n*phi
mu.n<-(1/phi.n)*(phi.0*m.0+n*phi*xbarra)
par_post <- cbind(mu.n,1/phi.n)

meus_par <- rbind(par_pri,par_vero,par_post)
meus_par <- as.data.frame(meus_par)
names(meus_par) <- c("média","desvio padrão")
row.names(meus_par) <- c("priori","verossimilhança","posteriori")
meus_par
```

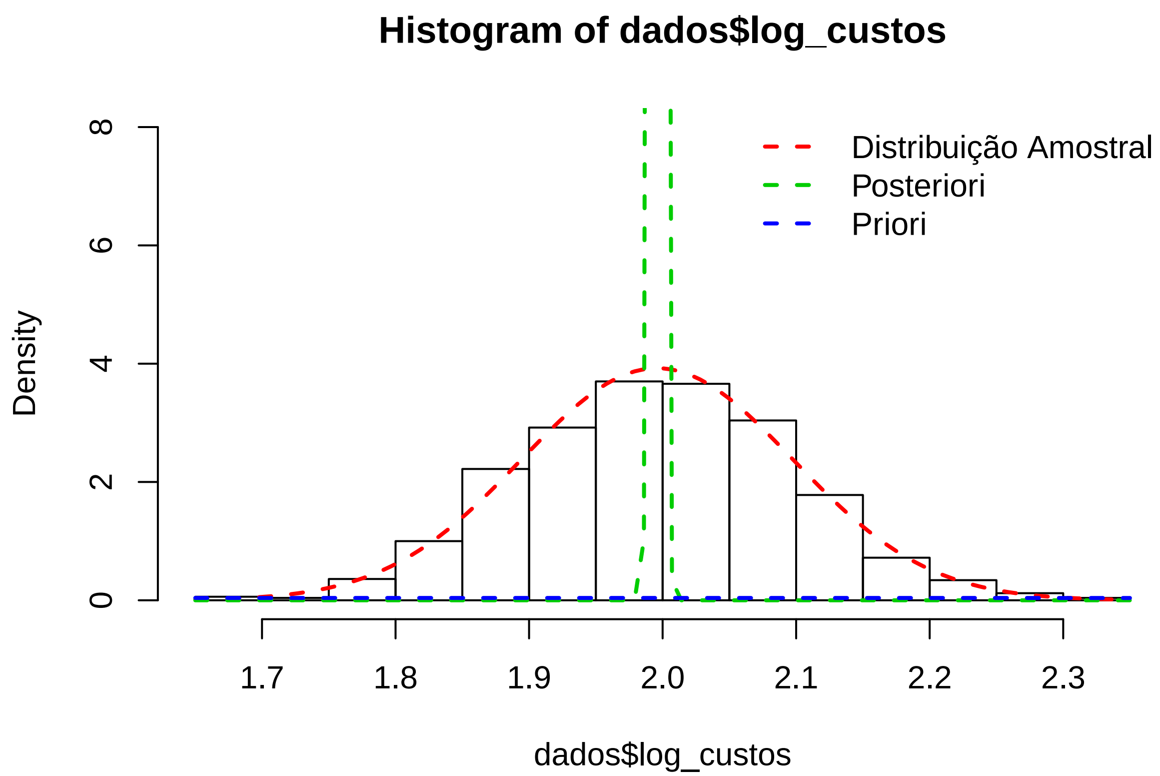
```
##                    média desvio padrão
## priori          0.000000  1.000000e+02
## verossimilhança 1.995955  1.033432e-02
## posteriori      1.995955  1.033432e-05
```

b) Gráfico do parâmetro a priori e posteriori

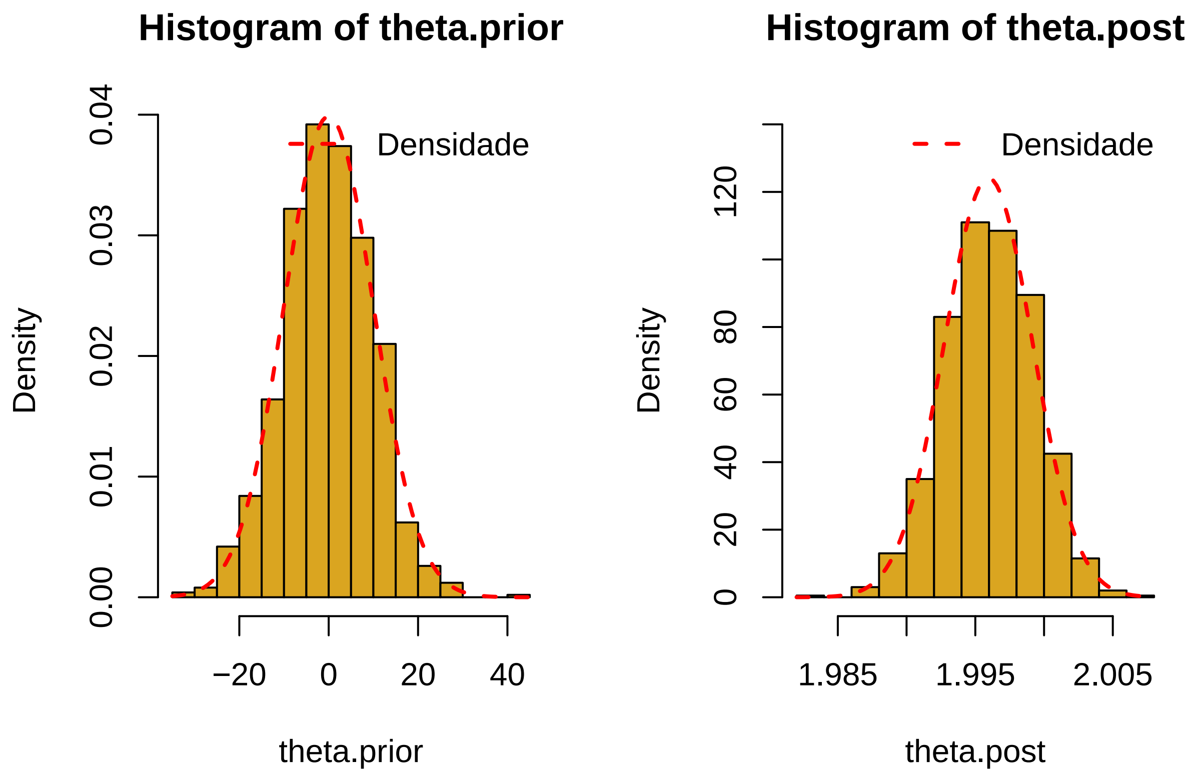
A amostra fez com que as crenças que tinhamos a priori fossem mudadas. Obtivemos uma média diferente da descrita a priori e uma variância muito pequena, muitas ordens de magnitude menor que o 100 especificado pela priori.

```
theta.post<- rnorm(n,mu.n,1/sqrt(phi.n))
theta.prior <- rnorm(n,m.0,1/sqrt(phi.0))

hist(dados$log_custos,prob=TRUE,ylim=c(0,8))
curve(dnorm(x,xbarra,sqrt(1/phi)),col=2,lty=2,lwd=2,add=TRUE)
curve(dnorm(x,mu.n,sqrt(1/phi.n)),col=3,lty=2,lwd=2,add=TRUE)
curve(dnorm(x,m.0,sqrt(1/phi.0)),col=4,lty=2,lwd=2,add=TRUE)
```

```
legend("topright",legend=c("Distribuição Amostral","Posteriori","Priori"),lty=c(2,2,2),lwd=2,
       col=c(2,3,4),bty="n")    # legenda
```




```
par(mfrow=c(1,2))
hist(theta.prior,prob=T,col="goldenrod")
curve(dnorm(x,m.0,1/sqrt(phi.0)),col=2,lty=2,lwd=2,add=TRUE)
legend("topright",legend=c("Densidade"),lty=c(2),lwd=2,
       col=c(2),bty="n")    # legenda

hist(theta.post,prob=T,col="goldenrod",ylim=c(0,140))
curve(dnorm(x,mu.n,1/sqrt(phi.n)),col=2,lty=2,lwd=2,add=TRUE)
legend("topright",legend=c("Densidade"),lty=c(2),lwd=2,
       col=c(2),bty="n")    # legenda
```



c) Estimativa intervalar, ao nível de credibilidade 95%, para a média dos log-custos a posteriori.

A média dos log-custos está entre 1.989204 e 2.001710, ao nível de credibilidade 95%.

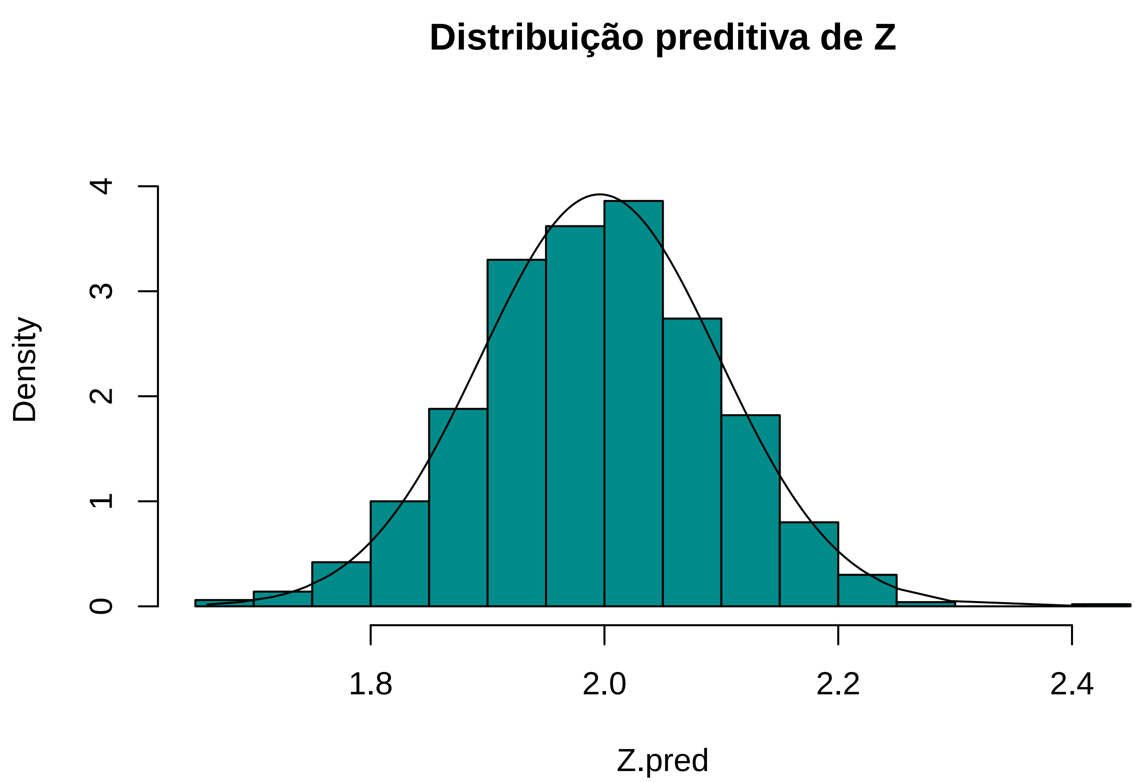
```
int.theta<-quantile(theta.post,probs=c(0.025,0.975))
int.theta
```

```
##      2.5%    97.5%
## 1.989454 2.002038
```

Etapa 2: Exploração do comportamento de uma observação futura e obtenção de distribuição preditiva.

d) Obtenha, por amostragem, uma aproximação para a distribuição preditiva de Z (logcusto). Faça um histograma da amostra da distribuição preditiva.

```
Z.pred<-NULL
for (i in 1:length(theta.post)){
  Z.pred[i]<-rnorm(1,theta.post[i],sqrt(1/phi))
}

hist(Z.pred,prob=T,col="darkcyan",main="Distribuição preditiva de Z",ylim=c(0,4.5))
phi.pred<-phi*phi.n/(phi.n+phi)
lines(sort(Z.pred),dnorm(sort(Z.pred),mu.n, sqrt(1/phi.pred)))
```



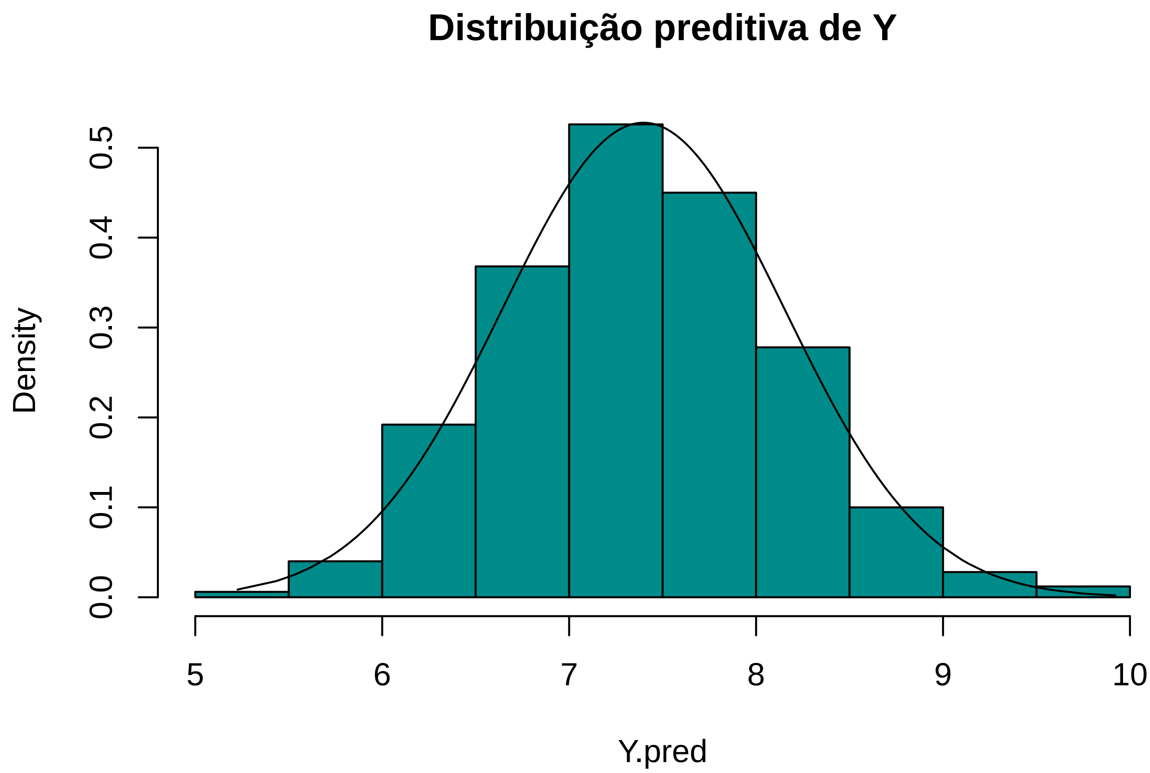
e) Aproximação para a distribuição preditiva de Y.

De forma semelhante a preditiva do log-custos, a verossimilhança tem um peso muito maior na distribuição a posteriori do que a distribuição a priori. Observamos que a preditiva também segue uma distribuição normal com a mesma média da verossimilhança.

```r
#Medidas amostrais
t2 <- sum(dados$custos)
phi2<-1/(sd(dados$custos))^2
xbarra2<-mean(dados$custos)
n2<-length(dados$custos)


#Priori -> theta~Normal(0,10^2):
m.0<-0
d.0<-10
phi.0<-1/d.0^2

#Posteriori:
phi2.n<-phi.0+n2*phi2
mu2.n<-(1/phi2.n)*(phi.0*m.0+n2*phi2*xbarra2)

#Precição
Y.pred<-NULL
for (i in 1:length(theta.post)){
  Y.pred[i]<-exp(rnorm(1,theta.post[i],sqrt(1/phi)))
}

#Parâmetros
par_post2 <- cbind(mu2.n,1/phi2.n)
par_pri2 <- cbind(m.0,1/phi.0)
par_vero2 <- cbind(xbarra2,1/phi2)
meus_par2 <- rbind(par_pri2,par_vero2,par_post2)
meus_par2 <- as.data.frame(meus_par2)
names(meus_par2) <- c("média","desvio padrão")
row.names(meus_par2) <- c("priori","verossimilhança","posteriori")
meus_par2
```

```
##                    média desvio padrão
## priori          0.000000  1.000000e+02
## verossimilhança 7.397366  5.707260e-01
## posteriori      7.397324  5.707227e-04
```

```r
hist(Y.pred,prob=T,col="darkcyan",main="Distribuição preditiva de Y")
phi2.pred<-phi2*phi2.n/(phi2.n+phi2)
lines(sort(Y.pred),dnorm(sort(Y.pred),mu2.n, sqrt(1/phi2.pred)))
```



f) Inferência sobre a preditiva de Y

```
# probabilidade de que um custo de cancelamento futuro ultrapasse 9 (mil Reais)
p<-length(Y.pred[Y.pred>=9])/length(Y.pred)
cat("A probabilidade de um custo de cancelamento futuro ultrapassar 9 mil reais é de", p, ".")
```

```
## A probabilidade de um custo de cancelamento futuro ultrapassar 9 mil reais é de 0.02 .
```

```
# o custo esperado de cancelamento;
Y.esp<-mean(Y.pred)
cat("O custo esperado de cancelamento é", Y.esp,".")
```

```
## O custo esperado de cancelamento é 7.396905 .
```

```
# estimativa intervalar
int.Y<-quantile(Y.pred,probs=c(0.025,0.975))
cat("A estimativa intervalar, ao nível de credibilidade 95%, fica entre", int.Y[1],"e",int.Y[2],".")
```

```
## A estimativa intervalar, ao nível de credibilidade 95%, fica entre 6.017238 e 8.920592 .
```