

Lista - distribuição amostral da média

Code ▾

Hide

```
library(data.table)
library(gtools)
```

Questão 1

Suponha que sua população seja formada pelos elementos $X = \{5;7;8;9\}$. a) Calcule a média populacional.

Hide

```
pop <- c(5,7,8,9)
mean(pop)
```

```
[1] 7.25
```

b. Calcule o desvio padrão populacional.

Hide

```
sqrt(((5-7.25)^2+(7-7.25)^2+(8-7.25)^2+(9-7.25)^2)/4)
```

```
[1] 1.47902
```

c. Construa o modelo/distribuição de probabilidades de X.

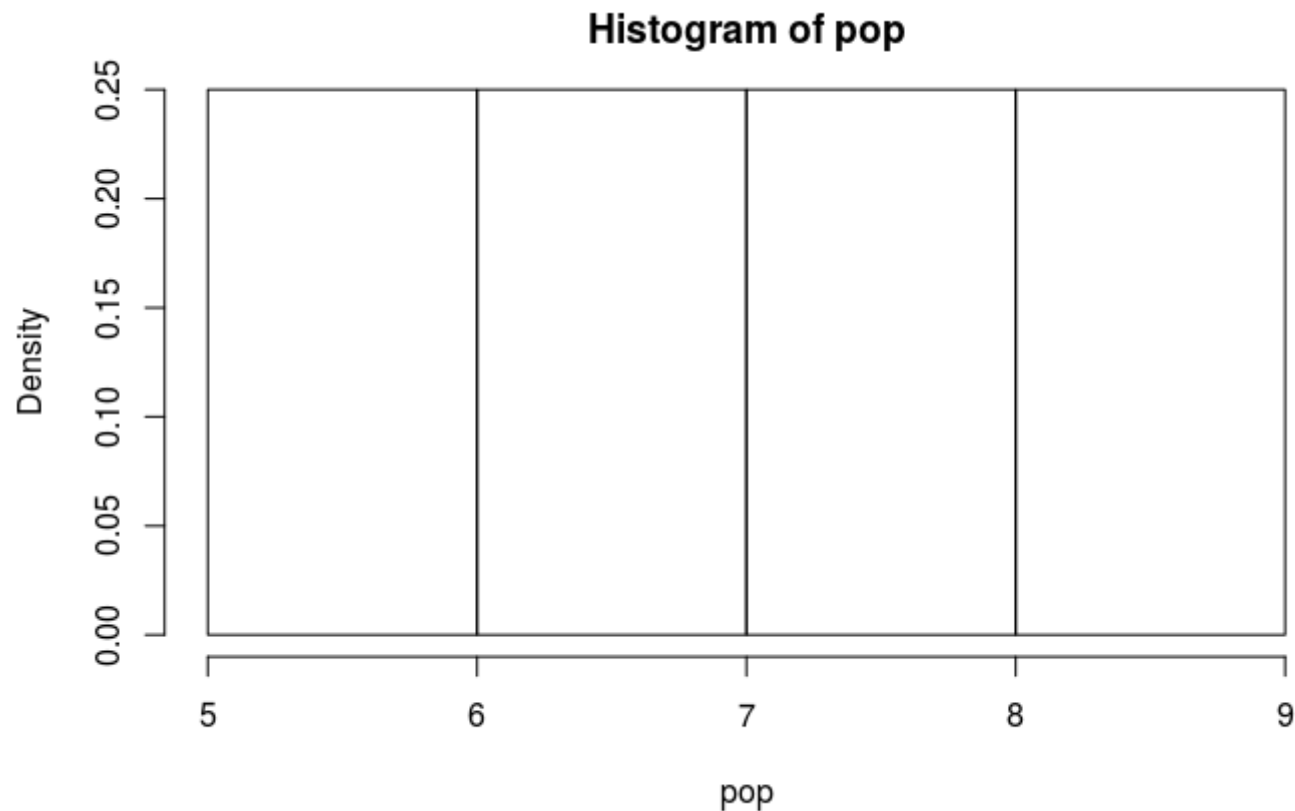
Hide

```
as.matrix(prop.table(table(pop)))
```

```
[,1]  
5 0.25  
7 0.25  
8 0.25  
9 0.25
```

[Hide](#)

```
hist(pop, freq=FALSE)
```



d. Construa o modelo de probabilidade das médias amostrais \bar{X} considerando amostras aleatórias de tamanho 3 com reposição.

[Hide](#)

```
amostras<-permutations(n=4,r=3,v=pop,repeats.allowed=T)
amostras <- rowMeans(amostras)
```

[Hide](#)

```
as.matrix(prop.table(table(amostras)))
```

```
      [,1]
5      0.015625
5.66666666666667 0.046875
6      0.046875
6.33333333333333 0.093750
6.66666666666667 0.093750
7      0.156250
7.33333333333333 0.140625
7.66666666666667 0.140625
8      0.109375
8.33333333333333 0.093750
8.66666666666667 0.046875
9      0.015625
```

e. Calcule a média das médias amostrais.

[Hide](#)

```
mean(amostras)
```

```
[1] 7.25
```

f. Calcule o desvio padrão das médias amostrais.

[Hide](#)

```
sqrt(sum((amostras-mean(amostras))^2)/length(amostras))
```

```
[1] 0.8539126
```

g. Explique o que é o modelo de probabilidade de X e o que é o modelo de probabilidade de \bar{X} .

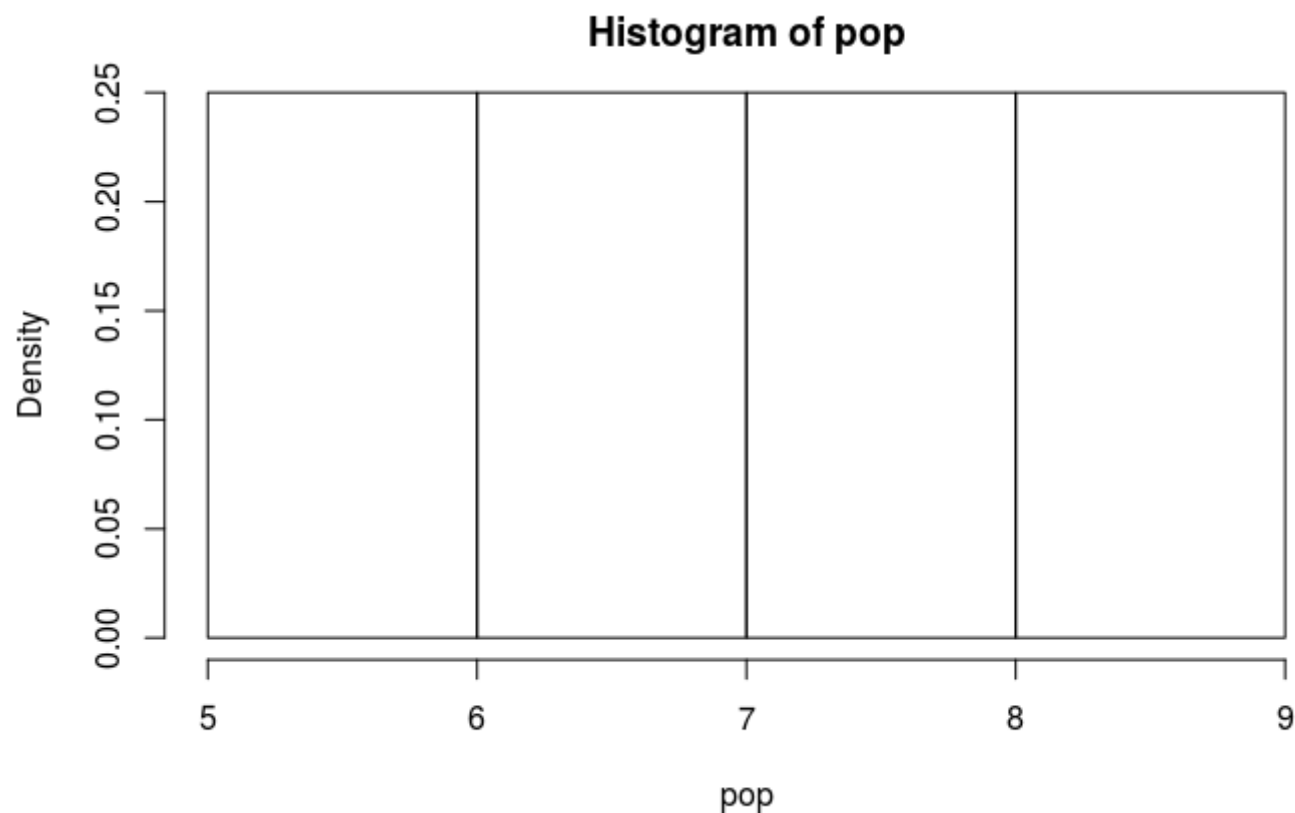
Modelo de probabilidade de X -> é a probabilidade de cada elemento da população ocorrer.

Modelo de probabilidade de \bar{X} -> é a probabilidade de cada média da distribuição de médias ocorrer.

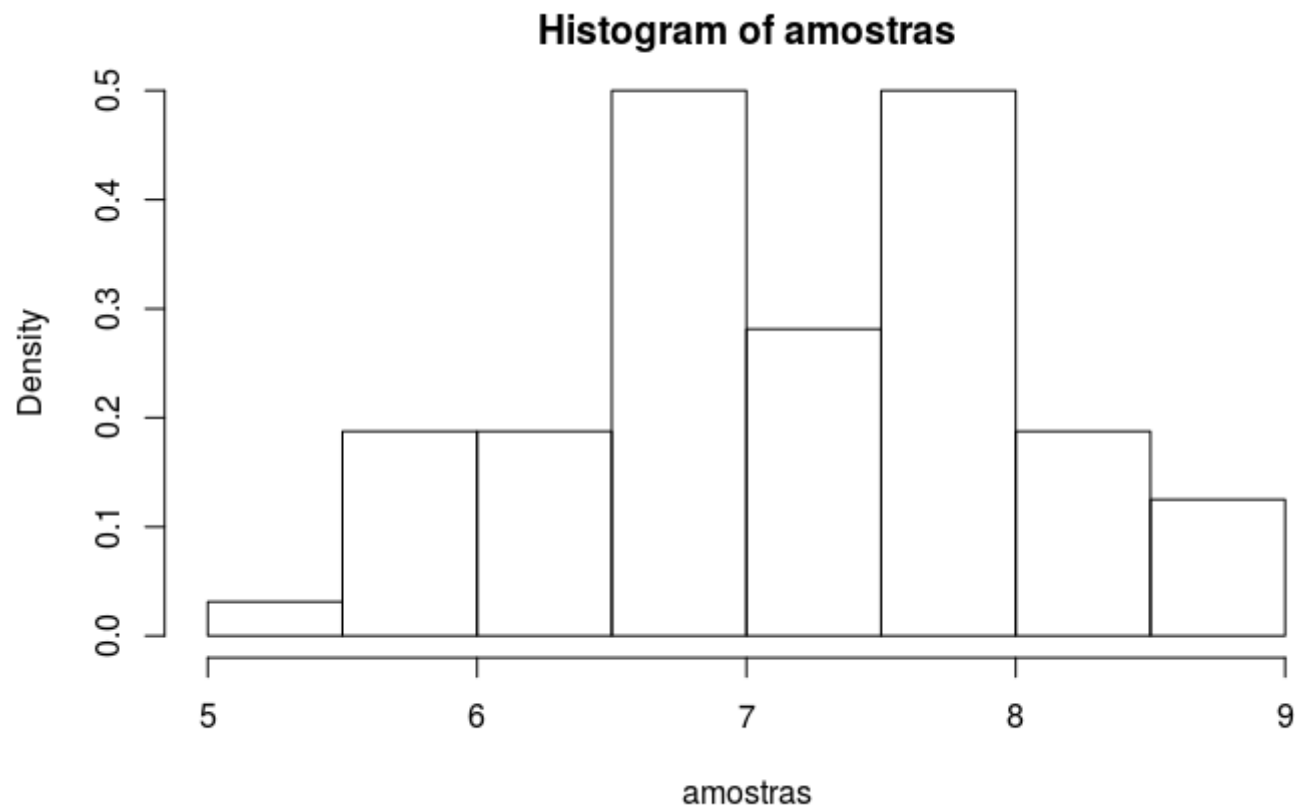
h. Faça os gráficos dos modelos de probabilidade de X e de \bar{X}

[Hide](#)

```
hist(pop, freq=FALSE)
```

[Hide](#)

```
hist(amostras, freq=FALSE)
```



Questão 2

Considere X uma variável aleatória de interesse com distribuição aproximadamente Normal com média 10 e desvio padrão 4. a) Qual a distribuição (e parâmetros) das médias amostrais considerando amostras independentes de tamanhos: $n = 10$; $n = 25$; $n = 100$; $n = 1000$.

[Hide](#)

```
set.seed(1)
pop<- rnorm(1000,10,4)
```

$n=10$

[Hide](#)

```
n=10
set.seed(1)
mediaAmostra = dpAmostra = NULL
for(i in 1:20000){
  Amostra <- sample(pop, n, replace = TRUE)
  mediaAmostra[i] <- mean(Amostra)
  dpAmostra[i] <- sd(Amostra) }
```

[Hide](#)

```
mean(mediaAmostra)
```

```
[1] 9.972347
```

[Hide](#)

```
sd(mediaAmostra)
```

```
[1] 1.311478
```

[Hide](#)

n=25

```
n=25
set.seed(1)
mediaAmostra = dpAmostra = NULL
for(i in 1:20000){
  Amostra <- sample(pop, n, replace = TRUE)
  mediaAmostra[i] <- mean(Amostra)
  dpAmostra[i] <- sd(Amostra) }
```

[Hide](#)

```
mean(mediaAmostra)
```

```
[1] 9.96091
```

[Hide](#)

```
sd(mediaAmostra)
```

```
[1] 0.8298169
```

n=100

[Hide](#)

```
n=100
set.seed(1)
mediaAmostra = dpAmostra = NULL
for(i in 1:20000){
  Amostra <- sample(pop, n, replace = TRUE)
  mediaAmostra[i] <- mean(Amostra)
  dpAmostra[i] <- sd(Amostra) }
```

[Hide](#)

```
mean(mediaAmostra)
```

```
[1] 9.95802
```

[Hide](#)

```
sd(mediaAmostra)
```

```
[1] 0.412499
```

n=1000

[Hide](#)

```
n=1000
set.seed(1)
mediaAmostra = dpAmostra = NULL
for(i in 1:20000){
  Amostra <- sample(pop, n, replace = TRUE)
  mediaAmostra[i] <- mean(Amostra)
  dpAmostra[i] <- sd(Amostra) }
```

Hide

```
mean(mediaAmostra)
```

```
[1] 9.953435
```

Hide

```
sd(mediaAmostra)
```

```
[1] 0.130555
```

- b. O que podemos notar a medida que o tamanho da amostra aumenta em relação a informação que temos à respeito da média da variável X?

A medida que o tamanho da amostra aumenta a média da variável X diminui.

Questão 3

Considere que o colesterol de mulheres saudáveis norte-americanas com idade entre 20 e 34 anos seja normalmente distribuído com média 186 mg/dL e desvio padrão 36 mg/dL. (Adaptado de National Center for Health Statistics) a) Qual a variável aleatória em questão? colesterol de mulheres saudáveis norte-americanas com idade entre 20 e 34 anos

- b. Qual a distribuição da variável aleatória? Supõe-se, que seja uma distribuição NORMAL com média 186 e desvio padrão 36 (rever)
- c. Qual a probabilidade de uma mulher vinda da população em questão, sorteada ao acaso, ter colesterol acima de 150 mg/dL? Interprete o valor obtido.

Hide


```
1-pnorm(150,186,36)
```

```
[1] 0.8413447
```

84% das mulheres americanas com idade entre 20 e 34 anos tem colesterol acima de 15 mg/dl.

- d. Se uma amostra aleatória de tamanho 50 é retirada da população em questão, qual a probabilidade da média da amostra estar entre 180.1 mg/dL e 191.9 mg/dL?

[Hide](#)

```
set.seed(1)
populacao<-rnorm(1000,186,36)
mediaAmostra = NULL
dpAmostra     = NULL
for(i in 1:20000){
  Amostra <- sample(populacao, 50, replace = TRUE)
  mediaAmostra[i] <- mean(Amostra)
  dpAmostra[i]    <- sd(Amostra) }
```

[Hide](#)

```
(m<-mean(mediaAmostra))
```

```
[1] 185.6336
```

[Hide](#)

```
(d<-sd(mediaAmostra))
```

```
[1] 5.276673
```

[Hide](#)

```
round(pnorm(191.9,m,d)-pnorm(180.1,m,d),digits=2)
```

```
[1] 0.74
```

e. Se uma amostra aleatória de tamanho 100 é retirada da população em questão, qual a probabilidade da média da amostra estar entre 180.1 mg/dL e 191.9 mg/dL?

Hide

```
set.seed(1)
populacao<-rnorm(1000,186,36)
mediaAmostra = NULL
dpAmostra     = NULL
for(i in 1:20000){
  Amostra <- sample(populacao, 100, replace = TRUE)
  mediaAmostra[i] <- mean(Amostra)
  dpAmostra[i]    <- sd(Amostra) }
```

Hide

```
(m<-mean(mediaAmostra))
```

```
[1] 185.6228
```

Hide

```
(d<-sd(mediaAmostra))
```

```
[1] 3.709943
```

Hide

```
round(pnorm(191.9,m,d)-pnorm(180.1,m,d),digits=2)
```

```
[1] 0.89
```

f. Qual a vantagem de se utilizar uma amostra de tamanho 100 em vez de uma amostra de tamanho 50?

Reduz-se o desvio padrão e as médias amostrais se aproximam da média da população.

A probabilidade das médias amostrais se afastarem da média populacional da variável X diminui

Questão 4

Supõe-se que o consumo mensal de água por residência em um certo bairro tenha distribuição Normal com média 10m³ e desvio padrão 2m³. Para uma amostra de 25 residências, escolhidas de forma aleatória, qual é a probabilidade de a média amostral não se afastar da verdadeira média populacional por mais de 1m³? O que aconteceria, em termos de probabilidade, se aumentássemos a amostra para 81 residências?

para n=25

[Hide](#)

```
set.seed(1)
populacao<-rnorm(1000,10,2)
mediaAmostra = NULL
dpAmostra     = NULL
for(i in 1:20000){
  Amostra <- sample(populacao, 25, replace = TRUE)
  mediaAmostra[i] <- mean(Amostra)
  dpAmostra[i]    <- sd(Amostra) }
```

[Hide](#)

```
(m<-mean(mediaAmostra))
```

```
[1] 9.980789
```

[Hide](#)

```
(d<-sd(mediaAmostra))
```

```
[1] 0.4145743
```

[Hide](#)

```
round(pnorm(11,m,d)-pnorm(9,m,d),digits=5)
```

```
[1] 0.98403
```

para n=81

[Hide](#)

```
set.seed(1)
populacao<-rnorm(1000,10,2)
mediaAmostra = NULL
dpAmostra     = NULL
for(i in 1:20000){
  Amostra <- sample(populacao, 81, replace = TRUE)
  mediaAmostra[i] <- mean(Amostra)
  dpAmostra[i]    <- sd(Amostra) }
```

[Hide](#)

```
(m<-mean(mediaAmostra))
```

```
[1] 9.979572
```

[Hide](#)

```
(d<-sd(mediaAmostra))
```

```
[1] 0.2293714
```

[Hide](#)

```
round(pnorm(11,m,d)-pnorm(9,m,d),digits=5)
```

```
[1] 0.99999
```