

Exercício EAD - 1ª Aula Washington

Raphael Hipólito

26/11/2019

Exercício 2

2.1 - Inserção dos dados no R

```
cod<- c(57, 60, 49, 50, 51, 60, 49, 53, 49, 56, 64, 60, 49, 52, 69, 40, 44, 38, 53, 66)
cod
```

```
## [1] 57 60 49 50 51 60 49 53 49 56 64 60 49 52 69 40 44 38 53 66
```

```
str(cod)
```

```
## num [1:20] 57 60 49 50 51 60 49 53 49 56 ...
```

Média

```
mean(cod)
```

```
## [1] 53.45
```

2.2 - Analise Exploratória

```
summary(cod)
```

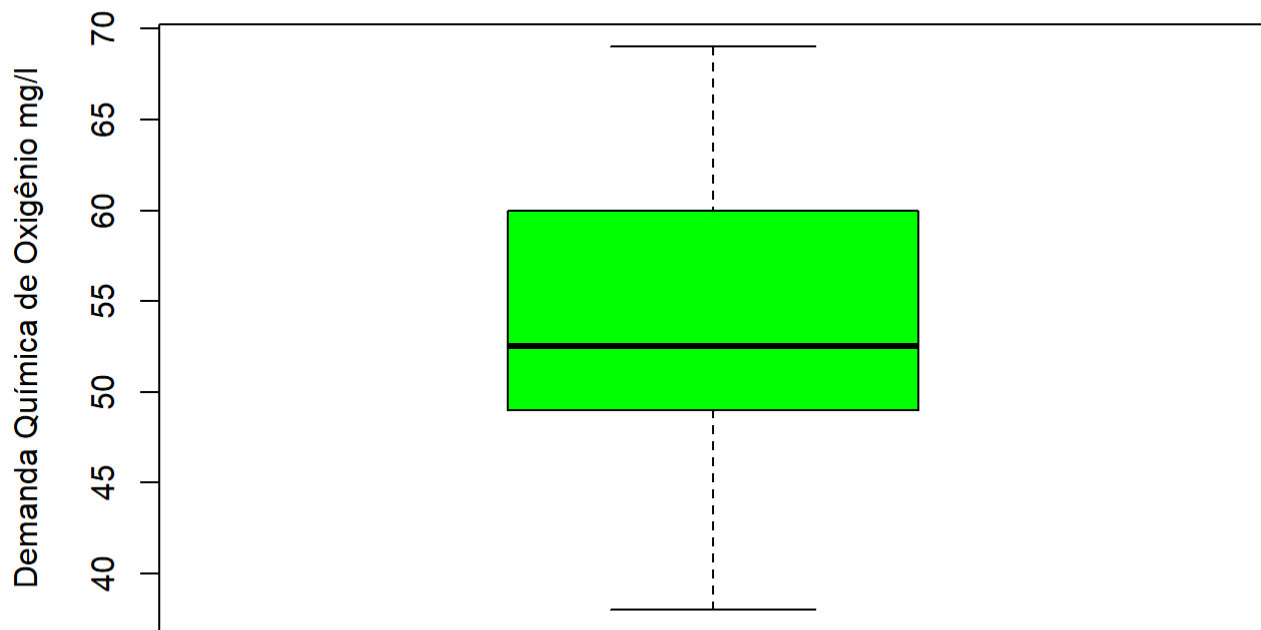
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  38.00   49.00   52.50   53.45   60.00   69.00
```

```
sort(cod)
```

```
## [1] 38 40 44 49 49 49 49 50 51 52 53 53 56 57 60 60 60 64 66 69
```

2.3 - Visualizando distribuição dos dados

```
boxplot(cod, col="green", ylab = "Demanda Química de Oxigênio mg/l")
```



2.4 - Teste de Shapiro-wilks de normalidade

```
shapiro.test(cod)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  cod  
## W = 0.97281, p-value = 0.8127
```

```
# Diferença observada  
mean(cod)
```

```
## [1] 53.45
```

```
### 4.1 - Teste de intervalo de confiança - Tstudent
```

```
#com alpha = 5%  
t.test(cod, alternative = c("two.sided"), mu = 50, conf.level = 0.95)
```

```
##  
## One Sample t-test  
##  
## data:  cod  
## t = 1.8894, df = 19, p-value = 0.07419  
## alternative hypothesis: true mean is not equal to 50  
## 95 percent confidence interval:  
##  49.62825 57.27175  
## sample estimates:  
## mean of x  
##      53.45
```

2.5 - Teste não-paramétrico de Wilcoxon

```
wilcox.test(cod, mu = 50, conf.int = TRUE)
```

```
##  
## Wilcoxon signed rank test with continuity correction  
##  
## data:  cod  
## V = 139, p-value = 0.07905  
## alternative hypothesis: true location is not equal to 50  
## 95 percent confidence interval:  
##  49.00001 58.00001  
## sample estimates:  
## (pseudo)median  
##      53.66572
```

Respostas as Questões - Exercício 2

A autorização de despejo para uma indústria requer que a concentração média mensal de COD seja inferior a 50 mg/L. A indústria quer que isso seja interpretado como “50 mg/L está dentro do intervalo de confiança da média, que vai ser estimada a partir de 20 observações por mês.”

a) Você concorda com a interpretação proposta pela indústria? Por que?

R: Não, a proposta da indústria é errônea. Mesmo que o valor da média da Demanda Química de Oxigênio ($n = 53,45$) esteja dentro da estimativa do intervalo de confiança (COD 49 a COD 58) pelo teste não-paramétrico de Wilcoxon e pelo Test T (COD 49,62 a COD 57,27), os valores medidos diariamente são absolutos e revelam um quantitativo superior ao autorizado. Assim, entende-se que a indústria deseja manipular erroneamente um modelo estatístico para que consiga despejar os efluentes. Baseando-se na rejeição da hipótese nula proposta pela indústria, podemos obter os valores de diferença observada em 7% (p-value) devido ao erro de amostragem aleatória.

b) Para as 20 observações seguintes, estaria a indústria em conformidade com a interpretação que você considerou adequada em a)?

R: A indústria continuaria despejando os efluentes acima do limite permitido pela sua autorização. Isto porque, o nível de significância é de 5%, sendo o valor de p-value (0.07905) superior, indicando que é provável que os dados de referência coletados sejam válidos para as 20 observações seguintes, rejeitando-se a hipótese nula e aceitando de que a concentração média mensal será superior ao permitido.

Exercício 3

Os seguintes dados foram obtidos a partir de medidas pareadas de nitrito em água e em águas residuais por eletrodo direto de íon-seletivo e um método colorimétrico.

3.1 - Inserção dos dados no R

```
ise = c(0.32, 0.36, 0.24, 0.11, 0.11, 0.44, 2.79, 2.99, 3.47)
cmetric = c(0.36, 0.37, 0.21, 0.09, 0.11, 0.42, 2.77, 2.91, 3.52)
ise
```

```
## [1] 0.32 0.36 0.24 0.11 0.11 0.44 2.79 2.99 3.47
```

```
cmetric
```

```
## [1] 0.36 0.37 0.21 0.09 0.11 0.42 2.77 2.91 3.52
```

```
str(ise)
```

```
##  num [1:9] 0.32 0.36 0.24 0.11 0.11 0.44 2.79 2.99 3.47
```

```
str(cmetric)
```

```
##  num [1:9] 0.36 0.37 0.21 0.09 0.11 0.42 2.77 2.91 3.52
```

3.2 - Análise exploratória

```
summary(ise)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.110    0.240    0.360    1.203    2.790    3.470
```

```
summary(cmetric)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.090    0.210    0.370    1.196    2.770    3.520
```

```
sort(ise)
```

```
## [1] 0.11 0.11 0.24 0.32 0.36 0.44 2.79 2.99 3.47
```

```
sort(cmetric)
```

```
## [1] 0.09 0.11 0.21 0.36 0.37 0.42 2.77 2.91 3.52
```

3.3 - Mediana e Análises Gerais

```
mean(ise)
```

```
## [1] 1.203333
```

```
mean(cmetric)
```

```
## [1] 1.195556
```

```
median(ise)
```

```
## [1] 0.36
```

```
median(cmetric)
```

```
## [1] 0.37
```

```
sd(ise)
```

```
## [1] 1.424798
```

```
sd(cmetric)
```

```
## [1] 1.421901
```

```
var(ise)
```

```
## [1] 2.03005
```

```
var(cmetric)
```

```
## [1] 2.021803
```

```
pnorm(ise, mean = 1.203333, sd = 1.424798, lower.tail = TRUE, log.p = FALSE)
```

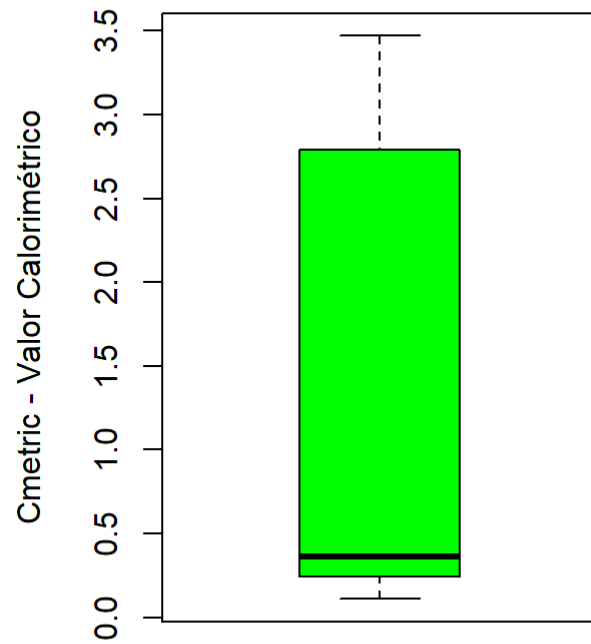
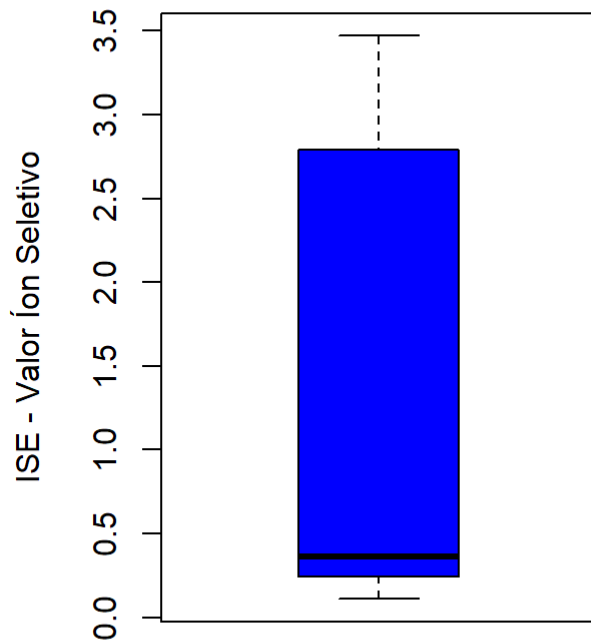
```
## [1] 0.2676385 0.2769599 0.2494825 0.2214338 0.2214338 0.2960663 0.8672764  
## [8] 0.8950752 0.9441805
```

```
pnorm(cmetric, mean = 1.195556, sd = 1.421901, lower.tail = TRUE, log.p = FALSE)
```

```
## [1] 0.2783893 0.2807550 0.2441154 0.2184261 0.2225963 0.2927268 0.8659138  
## [8] 0.8860413 0.9489486
```

3.4 - Gráfico Boxplot

```
par(mfrow=c(1,2))  
boxplot(ise, col="blue", ylab = "ISE - Valor Íon Seletivo")  
boxplot(cmetric, col="green", ylab = "Cmetric - Valor Calorimétrico")
```



3.5 - Teste de Shapiro-Wilks de Normalidade

```
shapiro.test(ise)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  ise  
## W = 0.72382, p-value = 0.00269
```

```
shapiro.test(cmetric)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  cmetric  
## W = 0.73167, p-value = 0.003321
```

3.6 - Teste T - Duas Amostras Independentes

```
t.test(ise, cmetric, paired = TRUE, var.equal = FALSE)
```

```
##  
## Paired t-test  
##  
## data: ise and cmetric  
## t = 0.5986, df = 8, p-value = 0.566  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.02218494 0.03774050  
## sample estimates:  
## mean of the differences  
## 0.007777778
```

3.7 - Teste não-paramétrico de Wilcoxon

```
wilcox.test(ise, cmetric, paired=TRUE, var.equal=FALSE)
```

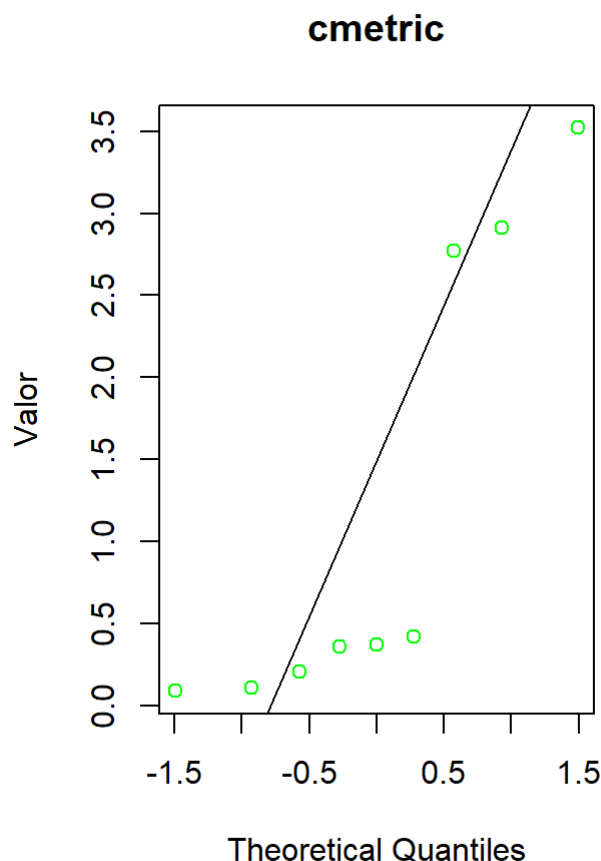
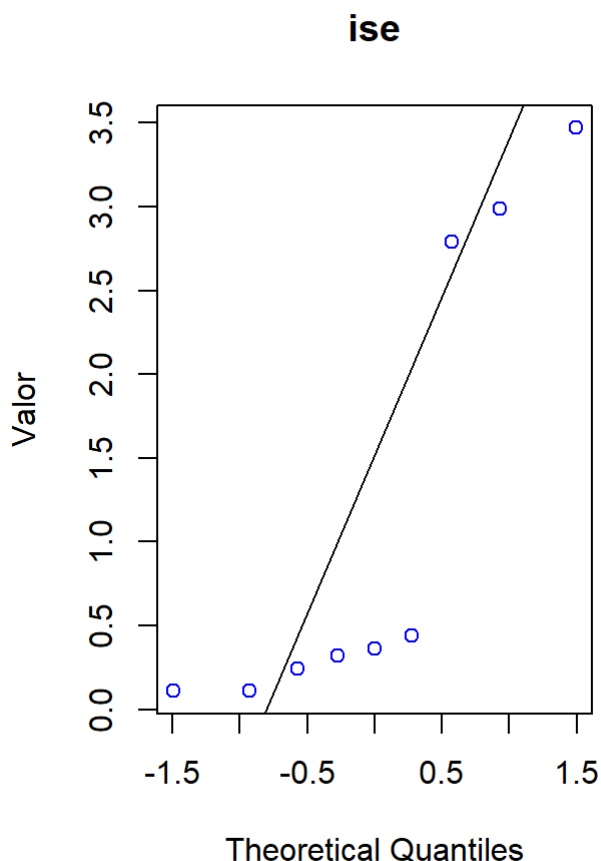
```
## Warning in wilcox.test.default(ise, cmetric, paired = TRUE, var.equal =  
## FALSE): cannot compute exact p-value with ties
```

```
## Warning in wilcox.test.default(ise, cmetric, paired = TRUE, var.equal =  
## FALSE): cannot compute exact p-value with zeroes
```

```
##  
## Wilcoxon signed rank test with continuity correction  
##  
## data: ise and cmetric  
## V = 22, p-value = 0.6236  
## alternative hypothesis: true location shift is not equal to 0
```

3.8 - Gráfico Teste Shapiro

```
par(mfrow=c(1,2))  
qqnorm(ise, main= "ise", ylab = "Valor", col = "blue")  
qqline(ise, col = "black")  
  
qqnorm(cmetric, main= "cmetric", ylab = "Valor", col = "green")  
qqline(cmetric, col = "black")
```



###

Com base nos dados disponibilizados, responda as seguintes questões:

a) Formule o problema em termos de um teste de hipótese e da estimação de um IC?

R: Tanto para os parâmetros ise e cmetric considera-se H_0 como hipótese nula e H_1 como Hipótese Alternativa.

O nível de confiança dos dados coletados é de 5%, sendo o valor de p-value (0.566 para Teste T e 0.6236 para Wilcox) superior ao nível de significância. Assim, rejeitamos a hipótese nula, pois os resultados indicam que é provável que os dados coletados sejam reais e a amostragem satisfatória. Para o teste de Shapiro-Wilks também rejeita-se a hipótese nula, pois por meio dos dados coletados, existe um indicativo que os dados estão fora da normalidade.

b) Os dois métodos forneceram resultados similares? Utilize um método tradicional adequado (paramétrico? Não-paramétrico?).

R: Para análise dos dados, foram utilizados testes paramétricos e não-paramétricos, observando-se em que ambos os testes os resultados encontrados foram similares. Devido a apresentação dos dados não seguirem uma distribuição normal, o método não-paramétrico é o mais indicado para apresentação dos dados por serem mais sensíveis aos erros de medidas do que os testes paramétricos.

Exercício 4

4.1 - a) Estime o poder do teste para os exercícios 2 e 3.

4.1.1 - Poder do Teste Exercício 2

```
library(pwr)
```

```
pwr.t.test(d = c(0.2, 0.5, 0.8), n = 20, sig.level = 0.05, type="one.sample", alternative="two.sided")
```



```
##
##      One-sample t test power calculation
##
##          n = 20
##          d = 0.2, 0.5, 0.8
##      sig.level = 0.05
##          power = 0.1359563, 0.5645044, 0.9238988
##      alternative = two.sided
```

4.1.2 - Poder do Teste Exercício 3

```
pwr.t.test(d = c (0.2, 0.5, 0.8), n = 9, sig.level = 0.05, type="one.sample", alternative="two.sided")
```

```
##
##      One-sample t test power calculation
##
##          n = 9
##          d = 0.2, 0.5, 0.8
##      sig.level = 0.05
##          power = 0.08291639, 0.26274609, 0.55909612
##      alternative = two.sided
```

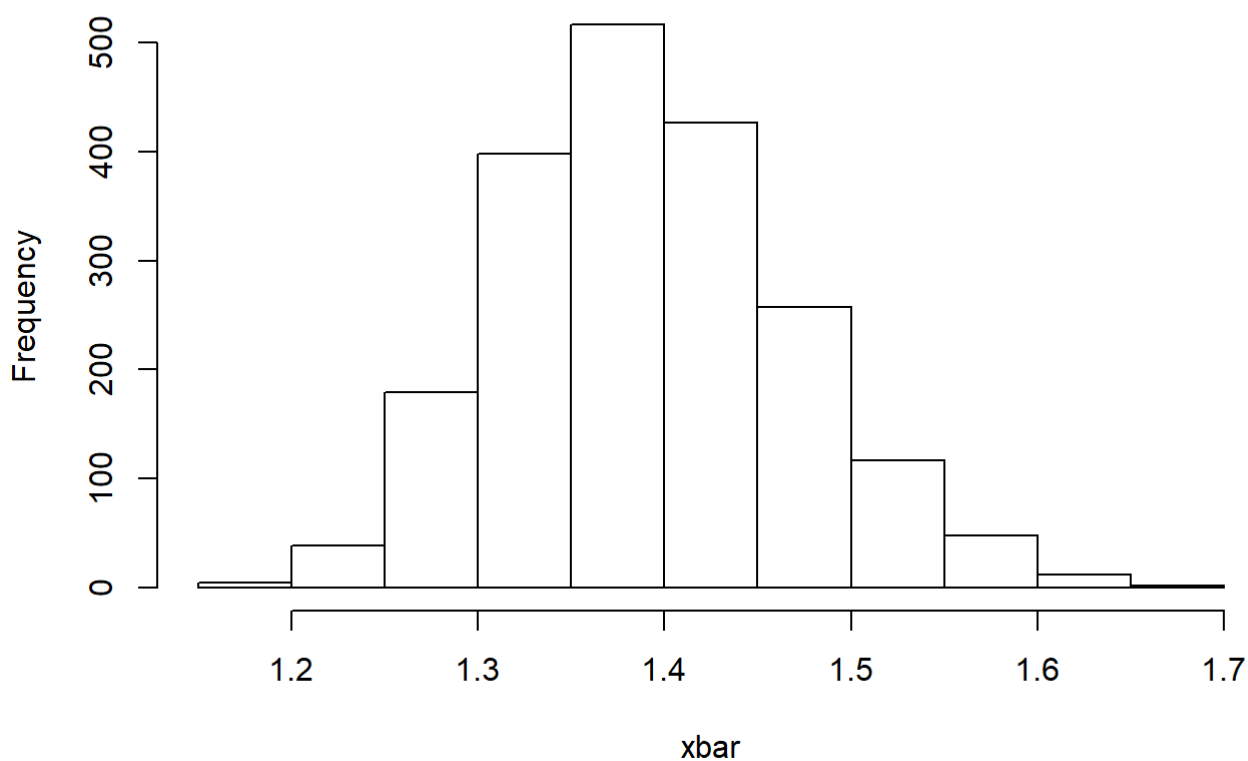
Exercício 5

5 - Obtenha IC via bootstrap para a média e para a mediana com os dados dos exercícios 1 e 2

5.1 - Exercício 1

```
od<-c(1.2, 1.4, 1.4, 1.3, 1.2, 1.35, 1.4, 2.0, 1.95, 1.1, 1.75, 1.05, 1.05, 1.4)
xbar= c()
for (i in 1:1999) {
  amostras = sample(od, size = length(od), replace = TRUE)
  xbar[i] = mean(amostras)
}
hist(xbar)
```

Histogram of xbar



```
quantile (xbar, c(.050, .950))
```

```
##          5%          95%  
## 1.275000 1.528929
```

5.1.2 - IC Para Média do Bootstrap

```
library(boot)  
# Dados Exercício 1  
od = c(1.2, 1.4, 1.4, 1.3, 1.2, 1.35, 1.4, 2.0, 1.95, 1.1, 1.75, 1.05, 1.05, 1.4)  
mean(od)
```

```
## [1] 1.396429
```

```
# IC Média Exercício 01  
fmedia = function(od, i) {mean(od[i])}  
boot.res1 = boot(data = od, statistic = fmedia, R = 1999)  
boot.ci(boot.res1, conf = 0.95, type = c("all"))
```

```
## Warning in boot.ci(boot.res1, conf = 0.95, type = c("all")): bootstrap  
## variances needed for studentized intervals
```

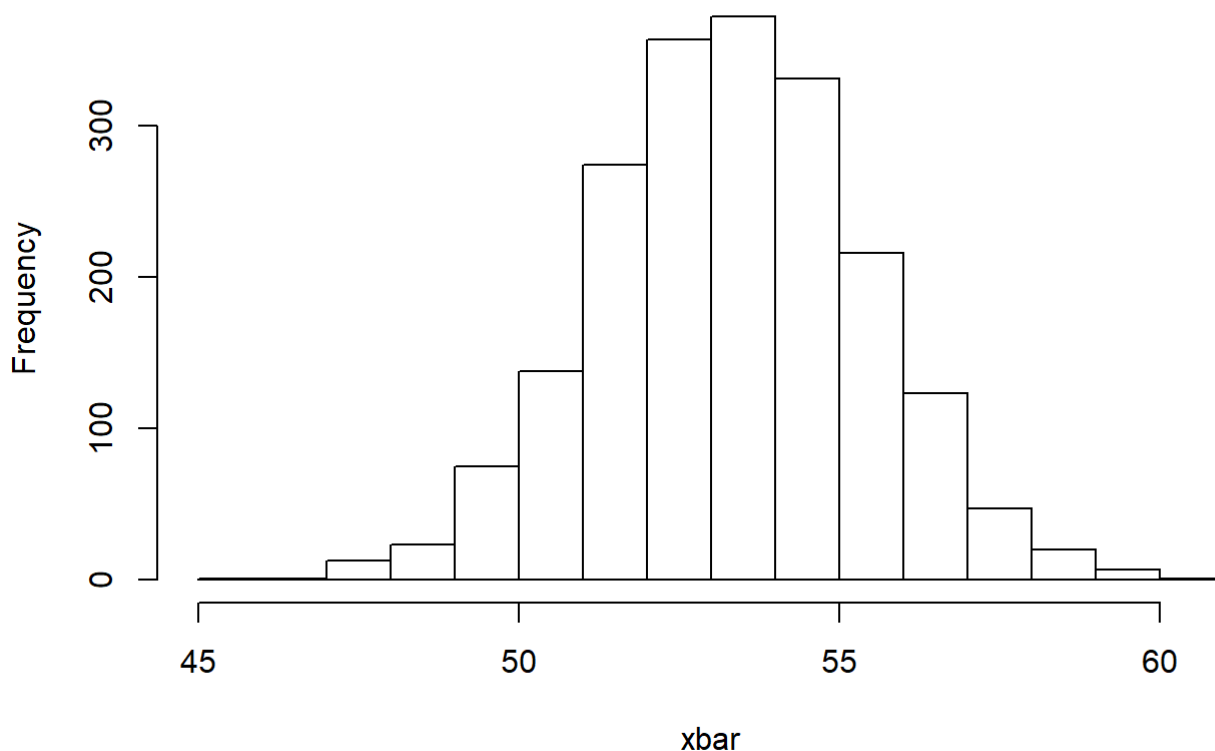
```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1999 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = boot.res1, conf = 0.95, type = c("all"))
##
## Intervals :
## Level      Normal          Basic
## 95%    ( 1.243,  1.549 )  ( 1.239,  1.543 )
##
## Level      Percentile      BCa
## 95%    ( 1.250,  1.554 )  ( 1.261,  1.564 )
## Calculations and Intervals on Original Scale
```

5.2 - Exercício 2

```
library(boot)
#Dados Exercício 2

cod<-c(57, 60, 49, 50, 51, 60, 49, 53, 49, 56, 64, 60, 49, 52, 69, 40, 44, 38, 53, 66)
xbar= c()
for (i in 1:1999) {
  amostras = sample(cod, size = length(od), replace = TRUE)
  xbar[i] = mean(amostras)
}
hist(xbar)
```

Histogram of xbar



```
quantile (xbar, c(.025, .975))
```

```
##      2.5%      97.5%
## 49.28571 57.42857
```

5.2.1 - IC Para Média do Bootstrap

```
library(boot)
# Dados Exercício 1
cod = c(57, 60, 49, 50, 51, 60, 49, 53, 49, 56, 64, 60, 49, 52, 69, 40, 44, 38, 53, 66)
mean(cod)
```

```
## [1] 53.45
```

```
# IC Média Exercício 01
fmedia = function(cod, i) {mean(cod[i])}
boot.res1 = boot(data = cod, statistic = fmedia, R = 1999)
boot.ci(boot.res1, conf = 0.95, type = c("all"))
```

```
## Warning in boot.ci(boot.res1, conf = 0.95, type = c("all")): bootstrap
## variances needed for studentized intervals
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1999 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = boot.res1, conf = 0.95, type = c("all"))
##
## Intervals :
## Level      Normal              Basic
## 95%   (49.98, 57.04 )   (49.80, 57.00 )
##
## Level      Percentile          BCa
## 95%   (49.90, 57.10 )   (49.95, 57.17 )
## Calculations and Intervals on Original Scale
```

...