EXAME_PARTE_II-MOQ-13

Bruno Lima Queiroz Santos

30/11/2019

Questão 1

a)

Um teste adequado a ser realizado é o teste t-student para as variações amostrais. Hipótese Nula: $H_0 : \mu = 4$, Hipótese alternativa: $H_a < 4$. Sendo a amostra de $n = 20$ elementos, com uma variância de $s^2 = 5.62 \Leftrightarrow s = \sqrt{5.62}$.

$$t_{n-1} \sim \frac{\overline{x} - \mu}{s/\sqrt{n}} \equiv \frac{\Delta x}{s/\sqrt{n}}$$

$$\Leftrightarrow P(\Delta x \geq 4) = P\left[t_{n-1} \geq \frac{4\sqrt{n}}{s}\right] = P\left[t_{n-1} \geq 7.54583\right]$$

```
n<-20
x<-4
x<-x*sqrt(20)
x<-x/sqrt(5.62)
cat("A probabilidade P(t_n-1>7.54583) = ", round(pt(x,n-1,lower.tail=FALSE),8))
```

```
## A probabilidade P(t_n-1>7.54583) =  2e-07
```

Criando uma amostra e testando-a.

```
X<-rnorm(n,mean=0,sd=sqrt(5.62))
Y<-X
Y<-Y-mean(Y)
Y<-Y/sd(Y)
Y<-Y*sqrt(length(Y))
t.test(Y,alternative="less",mu=4,conf.level=0.90)
```

```
##
##  One Sample t-test
##
## data:  Y
## t = -4, df = 19, p-value = 0.0003831
## alternative hypothesis: true mean is less than 4
## 90 percent confidence interval:
##      -Inf 1.327728
## sample estimates:
##    mean of x
## 3.469447e-17
```

De acordo com o teste realizado, a hipótese nula é descartada por um t-test de significância $\alpha = 0.10$ aplicado sobre uma amostra de distribuição normal com $\mu = 0$,$\sigma^2 = 5.62$

b)

O poder do teste é $P($ rejeitar $H_0|H_a$ verdadeira$)$. Assumindo que H_a é verdadeira, a probabilidade é justamente $P(\Delta x < 4) = P\left[t_{n-1} < \frac{4\sqrt{n}}{s} \equiv q\right]$

```r
quantil<-function(N,S,Mu){
  q<-Mu
  q<-q*sqrt(N)
  s<-sqrt(S) # STANDARD_DEVIATION by variance
  q<-q/s
  return(q)
}
q<-quantil(n,5.62, 4.0)
Power<-pt(q,n-1)
cat("O poder do teste é = ", Power)
```

```
## O poder do teste é =  0.9999998
```

O poder do teste é 99.99998%

Para $\sigma_A^2 = 7$:

```r
q<-quantil(n,7.0, 4.0)
Power<-pt(q,n-1)
cat("O poder do teste é = ", Power)
```

```
## O poder do teste é =  0.9999991
```

O poder do teste é 99.99991%

c)

$n \propto q^2 \ | \ P\Bigg[t < q\Bigg] = 0.90.$

```r
s<-sqrt(7.0)
n<-10
F_quantile<-function(n,s){
  Q<-(4.0*sqrt(n))/s
  return (p<-pt(Q,n-1))
}
p<-100
while(p>0.90 & n>1){
  p<-F_quantile(n,sqrt(7.0))
  n<-n-1
}
cat("n = ", n+1, "p = ", p)
```

```
## n =  2 p =  0.8607454
```

```r
F_quantile(2,sqrt(7.0))
```

```
## [1] 0.8607454
```

```r
F_quantile(3,sqrt(7.0))
```

```
## [1] 0.9399413
```

Portanto, é possível concluir que $n = 3$ garante potência de 0.90 com $\sigma_A^2 = 7$

Questão 2

```r
A<-c(776.58,672.00,1251.60,1133.58,797.58,965.58,735.00,1173.90,1238.58,1362.90,949.20,1385.58,839.58,5
B<-matrix(A,ncol=2)
B
```

```
##         [,1]    [,2]
## [1,]  776.58 1173.90
## [2,]  672.00 1238.58
## [3,] 1251.60 1362.90
## [4,] 1133.58  949.20
## [5,]  797.58 1385.58
## [6,]  965.58  839.58
## [7,]  735.00  562.38
```

```r
B<-B/4.20
A<-A/4.20
```

i) Preços em Dólar

```r
B
```

```
##       [,1]  [,2]
## [1,] 184.9 279.5
## [2,] 160.0 294.9
## [3,] 298.0 324.5
## [4,] 269.9 226.0
## [5,] 189.9 329.9
## [6,] 229.9 199.9
## [7,] 175.0 133.9
```

ii)

Média

```r
mean(A)
```

```
## [1] 235.4429
```

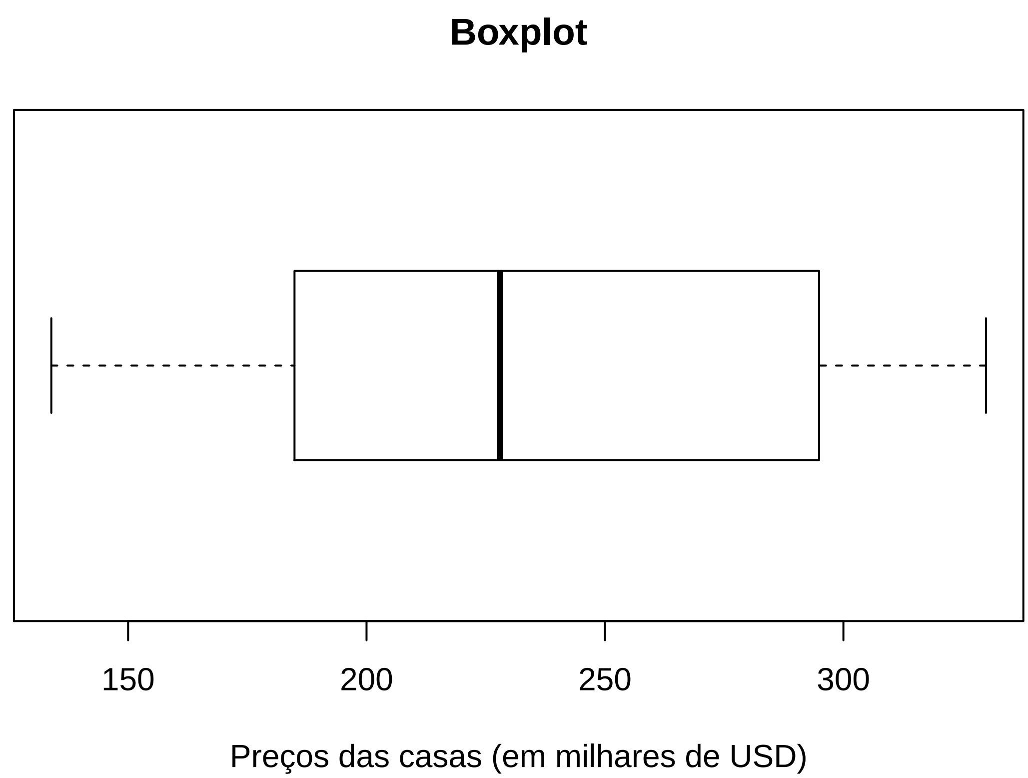
Desvio Padrão

```r
sd(A)
```

```
## [1] 63.90981
```

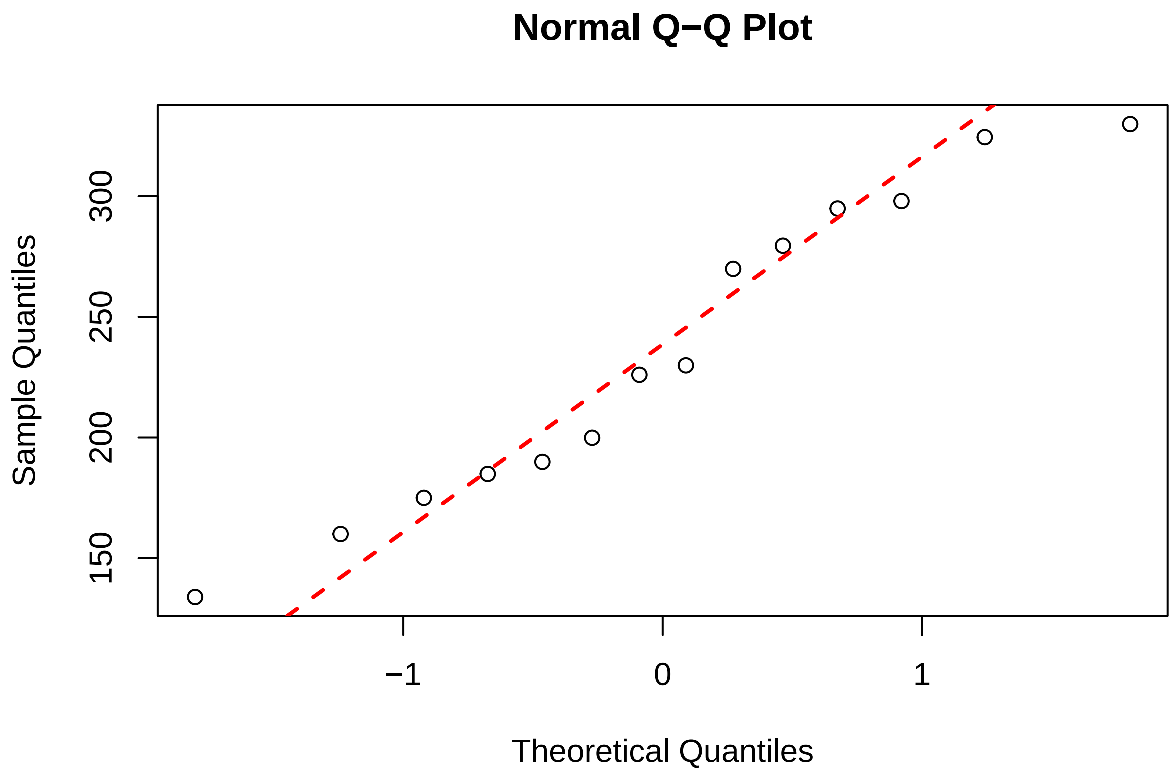
boxplot

```r
boxplot(
        A,
        xlab = "Preços das casas (em milhares de USD)",
        horizontal=TRUE,
        main = "Boxplot"
    )
```



QQPlot

```r
##qqnorm(A,xaxt="n",yaxt="n",xlab="",ylab="",main="Q-Q")
qqnorm(A)
qqline(A,col=2,lwd=2,lty=2)
```



É evidenciada uma distribuição simétrica em torno da média, ou de um valor razoavelmente próximo da média, levantando suspeitas sobre ser uma distribuição normal.

iii) Testes de normalidade

Teste de Kolmogorov-Smirnov [ks.test]

```
y<-rnorm(length(A))
ks.test(A,y)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  A and y
## D = 1, p-value = 4.985e-08
## alternative hypothesis: two-sided
```

```
y<-rnorm(length(A),sd=sd(A))
ks.test(A,y)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  A and y
## D = 1, p-value = 4.985e-08
## alternative hypothesis: two-sided
```

```r
y<-rnorm(length(A),mean=mean(A))
ks.test(A,y)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  A and y
## D = 0.57143, p-value = 0.01878
## alternative hypothesis: two-sided
```

```r
y<-rnorm(length(A),mean=mean(A),sd=sd(A))
ks.test(A,y)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  A and y
## D = 0.21429, p-value = 0.9205
## alternative hypothesis: two-sided
```

Teste de Shapiro-Wilk [shapiro.test]

```r
shapiro.test(A)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  A
## W = 0.94397, p-value = 0.4715
```

Teste de Anderson-Darlin [ad.test]

```r
library(nortest)
ad.test(A)
```

```
##
##  Anderson-Darling normality test
##
## data:  A
## A = 0.3219, p-value = 0.4909
```

Lilliefors [lillie.test]

```r
lillie.test(A)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  A
## D = 0.13951, p-value = 0.6495
```

Conclusão:

Todos os testes apresentam p-valor maior que 5.0%, inclusive o teste de Kolmogorov-Smirnov, em uma análise mais atenta, observando-se que há compatibilidade entre uma amostra normal cujos parâmetros populacionais têm valores próximos dos parâmetros estimados a partir da amostra analisada, embora o teste rejeite a hipótese nula quando não se fornecem os parâmetros populacionais compatíveis com o da amostra, isto é, os parâmetros padrão $\mu = 0, \sigma = 1$ não são compatíveis com os da amostra. Ou seja, há evidências de que a amostra segue uma distribuição normal não padronizada, pois a hipótese nula de que a função de distribuição é igual a uma função de distribuição normal não pode ser descartada com significância de 0.05 por qualquer um dos testes.

iv)

$$t_{n-1} \sim \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

$$P\left[t_{(\alpha/2,n-1)} \leq \frac{\overline{x} - \mu}{s/\sqrt{n}} \leq t_{(1-\alpha/2,n-1)}\right] = \alpha$$

$$\Leftrightarrow P\left[(s/\sqrt{n})t_{(\alpha/2,n-1)} \leq \overline{x} - \mu \leq t_{(1-\alpha/2,n-1)}(s/\sqrt{n})\right] = \alpha$$

$$\Leftrightarrow P\left[(-)t_{(1-\alpha/2,n-1)}(s/\sqrt{n}) \leq \mu - \overline{x} \leq (s/\sqrt{n})(-)t_{(\alpha/2,n-1)}\right] = \alpha$$

usando-se que :

$$0 = t_{(1-\alpha/2,n-1)} + t_{(\alpha/2,n-1)} \Leftrightarrow -t_{(\alpha/2,n-1)} = +t_{(1-\alpha/2,n-1)}$$

$$\Rightarrow P\left[\overline{x} - t_{(1-\alpha/2,n-1)}(s/\sqrt{n}) \leq \mu \leq \overline{x} + t_{(1-\alpha/2,n-1)}(s/\sqrt{n})\right] = \alpha$$

Logo, $\alpha = 0.05 \Leftrightarrow \frac{\alpha}{2} = 0.025$ e $n = 14$:

$$\overline{x} - t_{(0.975,13)}(s/\sqrt{14}) \leq \mu \leq \overline{x} + t_{(0.975,13)}(s/\sqrt{14})$$

Com : $\overline{x} = \sum \frac{x_i}{n}, s = \sqrt{\frac{(x_i - \overline{x})^2}{n-1}}$

```
x<-sum(A)
n<-length(A)
x<-x/n
s<-(A-x)^2
s<-sum(s)
s<-s/(n-1)
s<-sqrt(s)

x_inf<-x-qt(0.975,13)*(s/sqrt(14))
x_sup<-x+qt(0.975,13)*(s/sqrt(14))
cat("95% por cento IC:", x_inf,"  ",x_sup)
```

```
## 95% por cento IC: 198.5424    272.3433
```

```
t.test(A)
```

```
##
##  One Sample t-test
##
## data:  A
## t = 13.784, df = 13, p-value = 3.898e-09
## alternative hypothesis: true mean is not equal to 0
```

```
## 95 percent confidence interval:
##  198.5424 272.3433
## sample estimates:
## mean of x
##  235.4429
```

De onde obtêm-se os mesmos resultados para o intervalo de confidência 95%.

v)

```r
t.test(A,conf.level=0.95)
```

```
##
##  One Sample t-test
##
## data:  A
## t = 13.784, df = 13, p-value = 3.898e-09
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  198.5424 272.3433
## sample estimates:
## mean of x
##  235.4429
```

De acordo com o intervalo de confidência de 95%, há evidências com significância de 0.05 de que a alegação do corretor é falsa, pois a média alegada está fora do Intervalo de Confidência de 95%. Ou seja, a estimativa é de que a cada 100 amostras, apenas 5 contenham a média de preço fora do intervalo de confidência, ao contrário das demais 95.