

Estatística Aplicada I

Primeira Lista de Exercícios

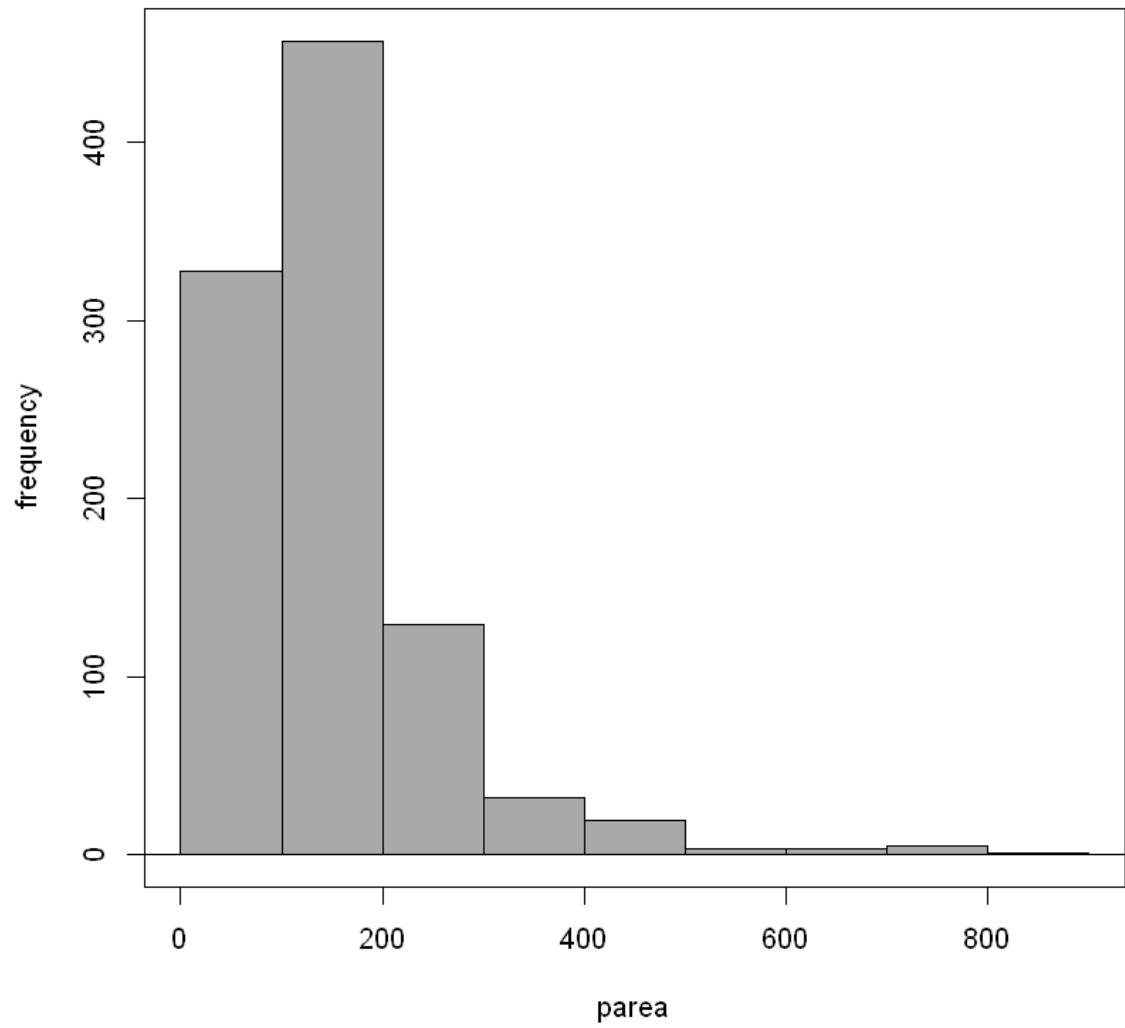
Estudante: Clístenes Grizafis Bento

Com a base de dados "realestateiaa" obter os seguintes resultados com o auxílio do "R"

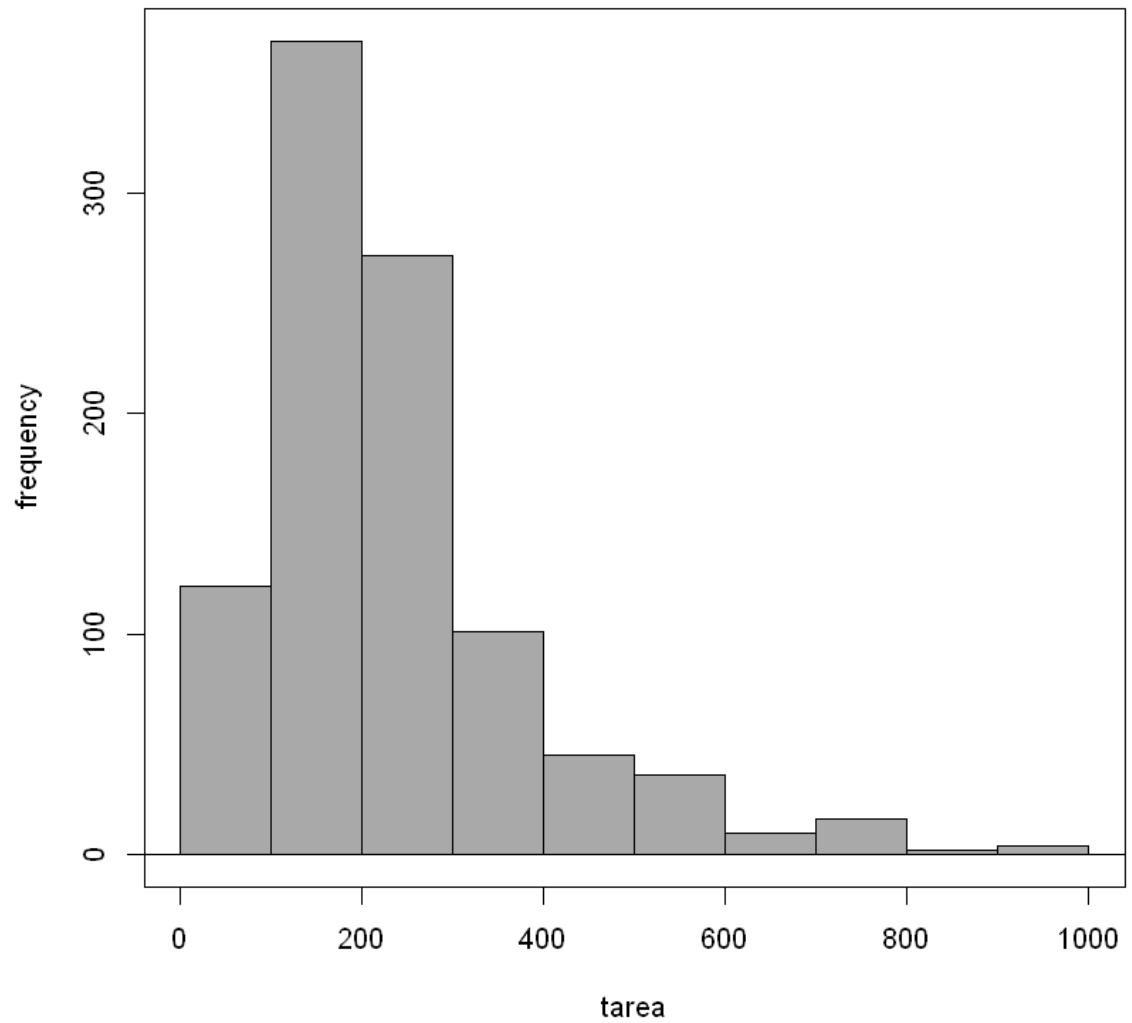
a) Elaborar o histograma e o boxplot das variáveis "parea e tarefa".

```
In [18]: #Carreando bibliotecas
load("realestateiaa.RData")
library (RcmdrMisc)
library (fdth)
library(BSDA)
library(nortest)
```

```
In [19]: #Histograma da variável "parea"
with(realestateiaa, Hist(parea, scale="frequency", breaks="sturges", col="darkgray")
```

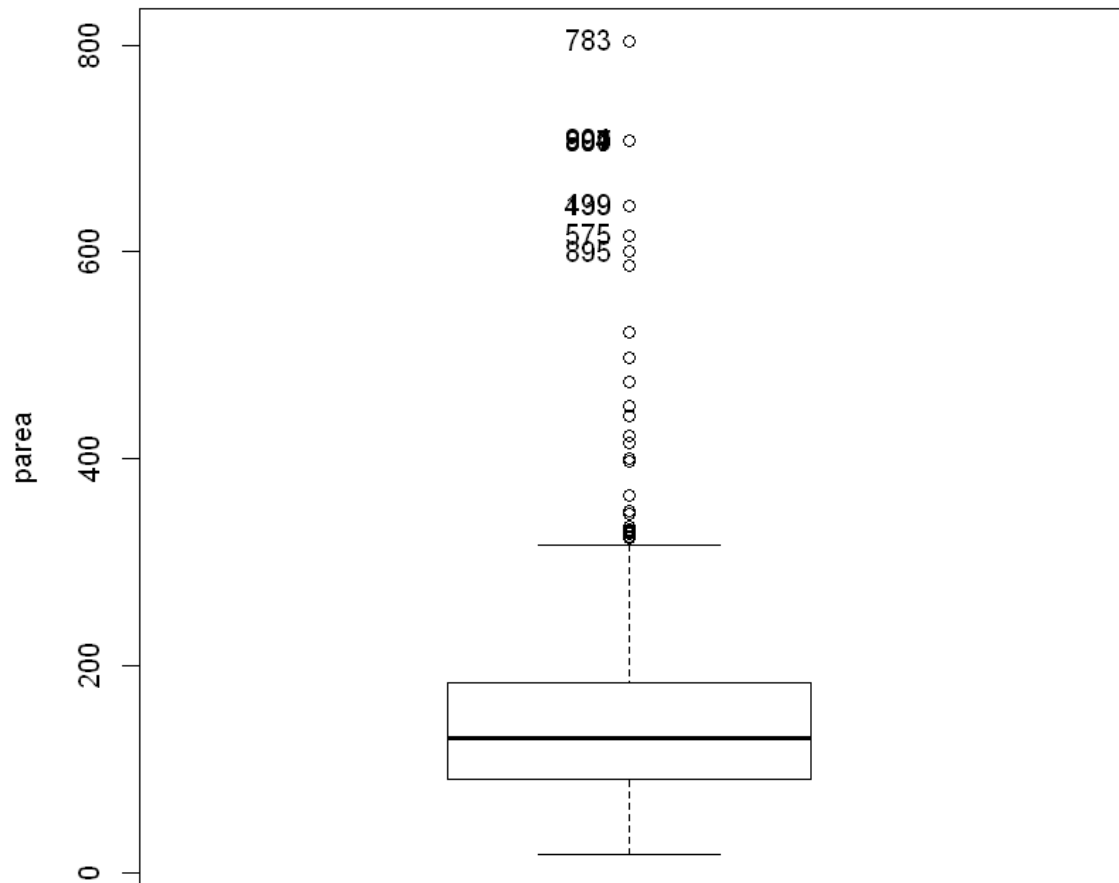


```
In [21]: # Histograma da variável "tarea"
with(realstateiaa, Hist(tarea, scale="frequency", breaks="sturges", col="darkgray"
```



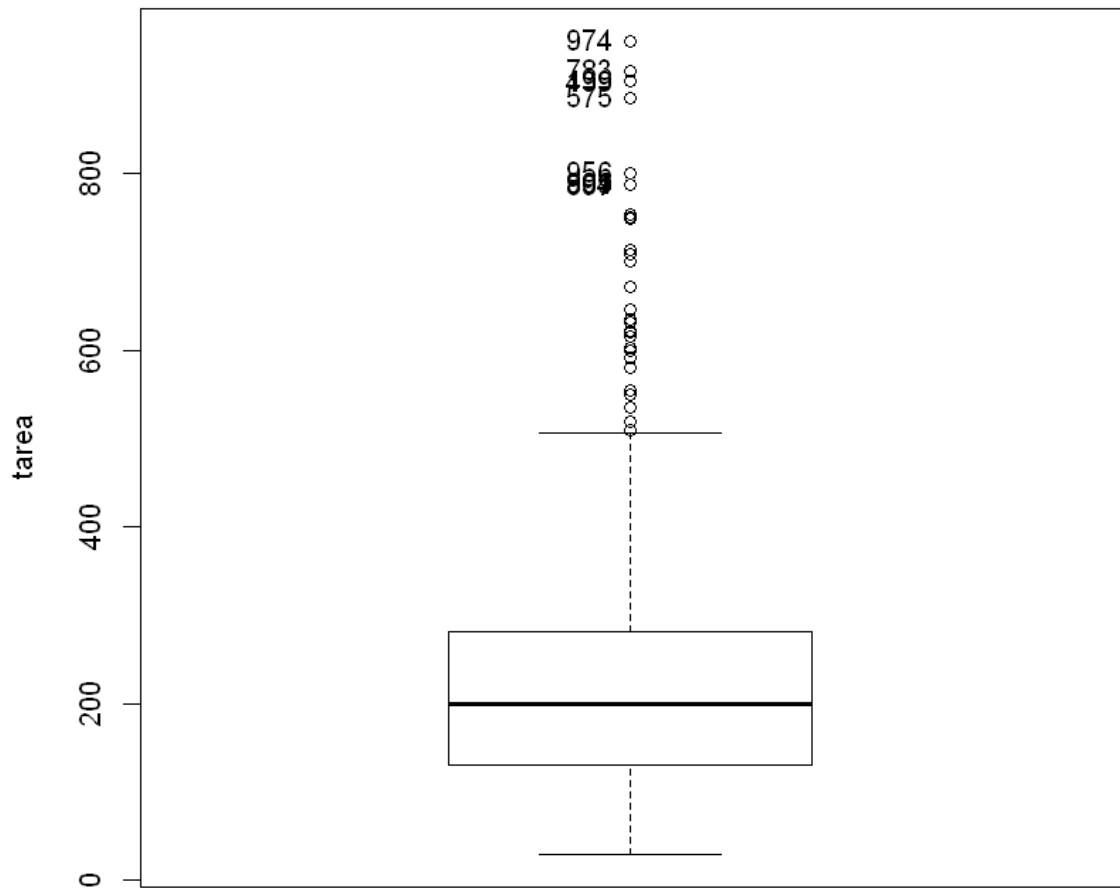
```
In [22]: # Bloxpot da variável "parea"  
Boxplot( ~ parea, data=realestateiaa, id=list(method="y"))
```

1. '783'
2. '891'
3. '904'
4. '897'
5. '901'
6. '909'
7. '139'
8. '499'
9. '575'
10. '895'



```
In [23]: # Bloxpot da variável "tarea"
Boxplot( ~ tarefa, data=realestateiaa, id=list(method="y"))
```

1. '974'
2. '783'
3. '139'
4. '499'
5. '575'
6. '956'
7. '891'
8. '904'
9. '909'
10. '897'



b) Elaborar a tabela de distribuição de frequências da variável "price" (preço dos imóveis)

```
In [36]: df <- fdt (realestateiaa$price)$table
df
```

Class limits	f	rf	rf(%)	cf	cf(%)
[157410,591062.5)	279	0.285568066	28.5568066	279	28.55681
[591062.5,1024715)	288	0.294779939	29.4779939	567	58.03480
[1024715,1458368)	162	0.165813715	16.5813715	729	74.61617
[1458368,1892020)	117	0.119754350	11.9754350	846	86.59161
[1892020,2325673)	53	0.054247697	5.4247697	899	92.01638
[2325673,2759325)	28	0.028659161	2.8659161	927	94.88229
[2759325,3192978)	17	0.017400205	1.7400205	944	96.62231
[3192978,3626630)	8	0.008188332	0.8188332	952	97.44115
[3626630,4060283)	14	0.014329580	1.4329580	966	98.87410
[4060283,4493935)	8	0.008188332	0.8188332	974	99.69294
[4493935,4927588)	3	0.003070624	0.3070624	977	100.00000

Legenda: rf - Frequência Relativa, cf - Frequência Acumulada

c) Para a variável “price” calcular os seguintes indicadores: média; mediana; moda; variância; desvio padrão; CV– Coeficiente de Variação; Quartis; distância interquartílica; percentis.

```
In [46]: options(scipen = 999)
media = round(mean(realestateiaa$price), 2)
mediana = round(median(realestateiaa$price), 2)
moda = subset(table(realestateiaa$price), table(realestateiaa$price) == max(table(realestateiaa$price)))
modaVal = names(moda)
var = round(var(realestateiaa$price), 2)
dp = round(sd (realestateiaa$price), 2)
cv = round((sd(realestateiaa$price) / mean(realestateiaa$price)) * 100, 2)
q1 = quantile(realestateiaa$price, probs = 0.25)
q2 = quantile(realestateiaa$price, probs = 0.50)
q3 = quantile(realestateiaa$price, probs = 0.75)
di = round(IQR(realestateiaa$price), 2)
quant = quantile(realestateiaa$price, c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0))
```

```
In [83]: df2 <- data.frame(media=media, mediana=mediana, moda=moda, modaVal=modaVal, var=var, dp=dp, cv=cv, q1=q1, q2=q2, q3=q3, di=di, quant=quant)
df3 <- data.frame(quant=quant)
df4 <- t(df2)
names(df4) <- NULL
names(df3) <- NULL
df4
df3
```

	850000
media	1140123
mediana	950000
moda	16
modaVal	850000
var	670486776668
dp	818832.6
cv	71.82
q1	550000
q2	950000
q3	1485000
di	935000
10%	353000.0
20%	490000.0
30%	607056.8
40%	799400.0
50%	950000.0
60%	1090000.0
70%	1366000.0
80%	1678800.0
90%	2123200.0

d) Estimar o intervalo de confiança para a média da variável "price" com 95% de confiança

```
In [84]: z.test(realestateiaa$price, y = NULL, alternative = "two.sided", mu = 0, sigma.x =
          sigma.y = NULL, conf.level = 0.95)
```

One-sample z-Test

```
data: realestateiaa$price
z = 43.522, p-value < 0.00000000000000022
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 1088778 1191468
sample estimates:
mean of x
 1140123
```

Avaliação do resultado: A média de valores em "price" varia entre \$1088778 e \$1191468, com média de \$1140123, com 95% de confiança ou 5% de significância. Estatística z = 43.522, ou seja, é maior que a estatística tabelada, de 1.96, rejeitando-se a hipótese (H0) de que o valor verdadeiro da média é estatisticamente igual a zero.

e) Fazer o teste de diferença entre médias para as variáveis

“parea” e “tarea”.

```
In [86]: z.test(realestateiaa$parea, realestateiaa$tarea, alternative = "two.sided", mu = 0,  
              sigma.x = sd(realestateiaa$parea), sigma.y = sd(realestateiaa$tarea), conf. = 0.95)
```

Two-sample z-Test

```
data: realestateiaa$parea and realestateiaa$tarea  
z = -14.151, p-value < 0.00000000000000022  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 -94.09729 -71.20261  
sample estimates:  
mean of x mean of y  
 152.9867  235.6366
```

Avaliação do resultado: A diferença entre as médias estará entre -94.09729 e -71.20261, para as médias de 152.9867 e 235.6366 (diferença de -82,6499), com 95% de confiança ou 5% de significância. Estatística $z = -14.151$, ou seja, é menor que a estatística tabelada, de -1.96, rejeitando-se a hipótese (H_0) de que a diferença verdadeira entre as médias não é igual a zero. Logo, pode-se dizer que as médias são estatisticamente diferentes.

f) Fazer o teste de diferença entre variâncias para as variáveis “parea” e “tarea”.

```
In [87]: var.test(realestateiaa$parea, realestateiaa$tarea, alternative = "two.sided", conf = 0.95)
```

F test to compare two variances

```
data: realestateiaa$parea and realestateiaa$tarea  
F = 0.43855, num df = 976, denom df = 976, p-value <  
0.00000000000000022  
alternative hypothesis: true ratio of variances is not equal to 1  
95 percent confidence interval:  
 0.3868074 0.4972143  
sample estimates:  
ratio of variances  
 0.4385501
```

Avaliação do resultado: A razão de variância é de 0.4385501 e a estatística $F = 0.43855$ com 976 graus de liberdade no numerador e denominador. Como o valor de F não está entre os valores tabelados 0.900007 e 1.111102 (encontrados pela função $qf(0.95, 976, 976)$ e $1/qf(0.95, 976, 976)$), as variâncias não são estatisticamente iguais.

g) Fazer o Teste de Wilcoxon-Mann-Whitney para amostras independentes para as variáveis “parea” e “tarea”.

```
In [88]: wilcox.test(realestateiaa$parea, realestateiaa$tarea, alternative = "two.sided")
```

Wilcoxon rank sum test with continuity correction

```
data: realestateiaa$parea and realestateiaa$tarea  
W = 287419, p-value < 0.00000000000000022  
alternative hypothesis: true location shift is not equal to 0
```

Avaliação do resultado: Como o resultado p-value é menor que 0.05 as amostras são independentes.

h) Fazer 2 testes de normalidade (a sua escolha) para a variável "price".

Anderson-Darling

```
In [90]: ad.test(realestateiaa$price)
```

Anderson-Darling normality test

data: realestateiaa\$price
A = 35.862, p-value < 0.0000000000000022

Avaliação do resultado: Como p-value é menor que 0.05 rejeita-se a hipótese de normalidade da variável

Kolmogorov-Smirnov

```
In [89]: normalityTest(~price, test="lillie.test", data=realestateiaa)
```

Lilliefors (Kolmogorov-Smirnov) normality test

data: price
D = 0.14486, p-value < 0.0000000000000022

Avaliação do resultado: Valor calculado de D = 0.14486 e p-value menor que 0.05, rejeita-se a hipótese de normalidade da amostra.