

# Estatística 1

## Primeira lista de exercício

a) Elaborar o histograma e o boxplot das variáveis “parea e tarea”.

Nos gráficos de frequência abaixo nota-se que existem muitos outliers tanto em parea quanto em tarea.

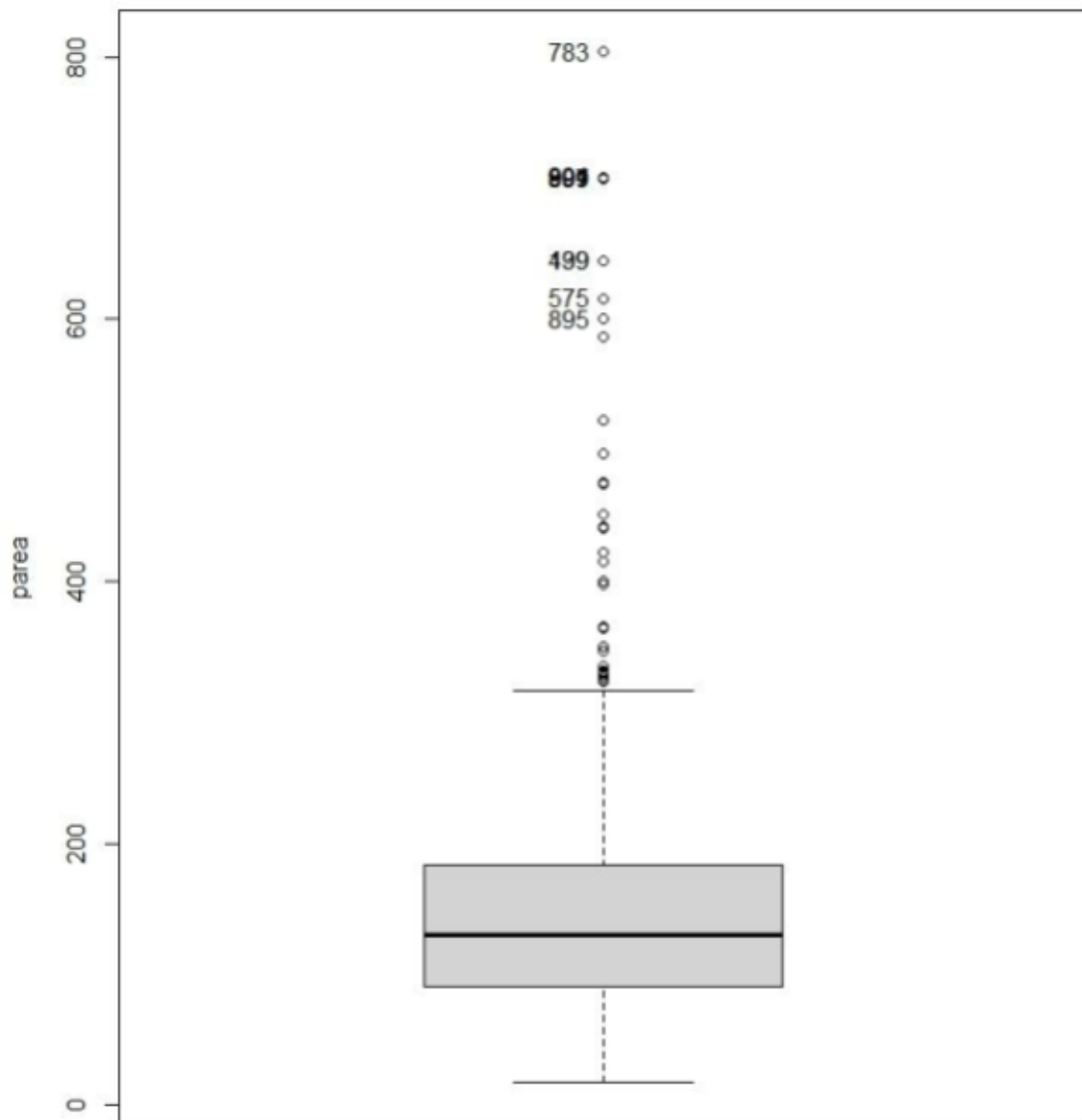
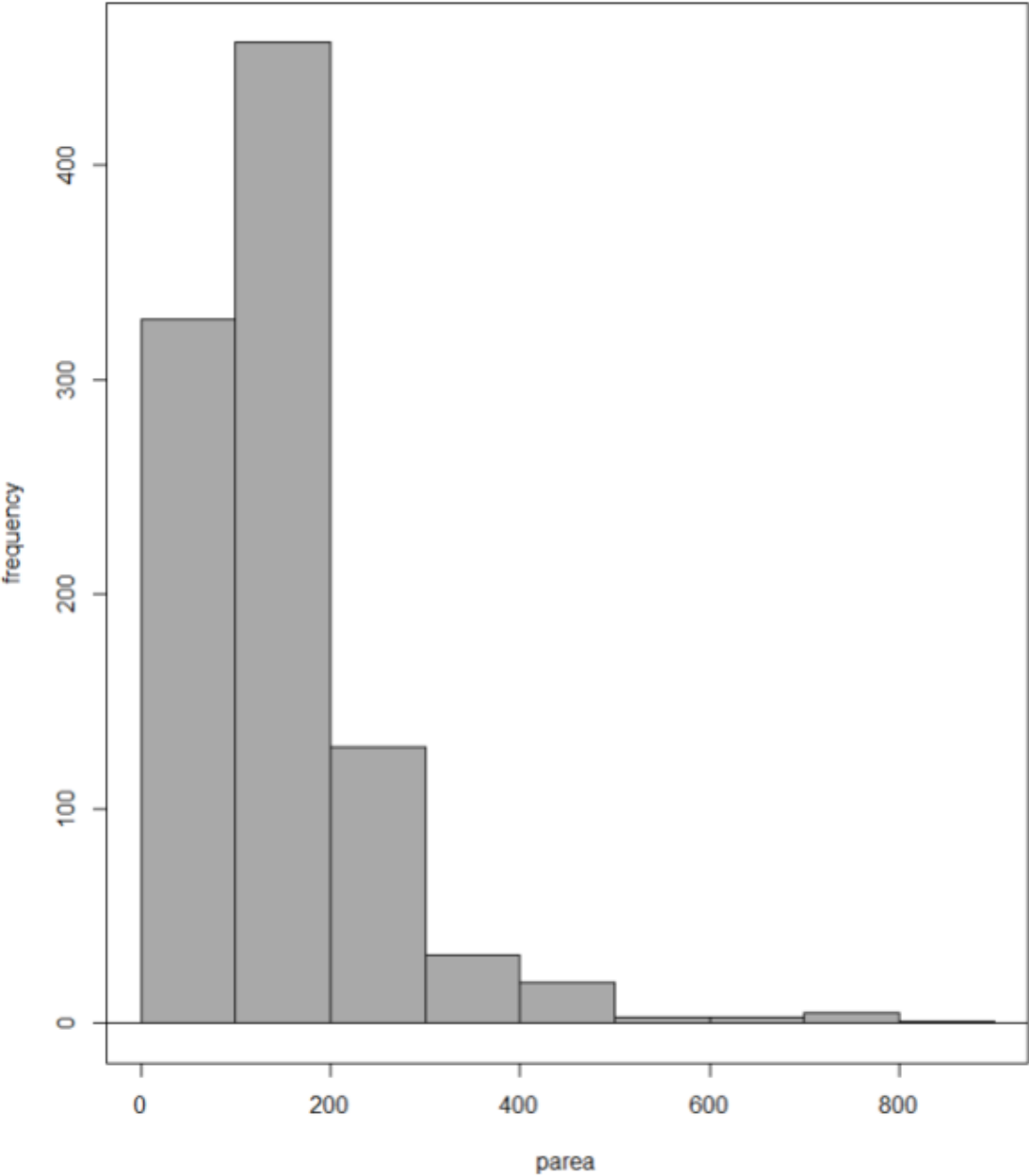
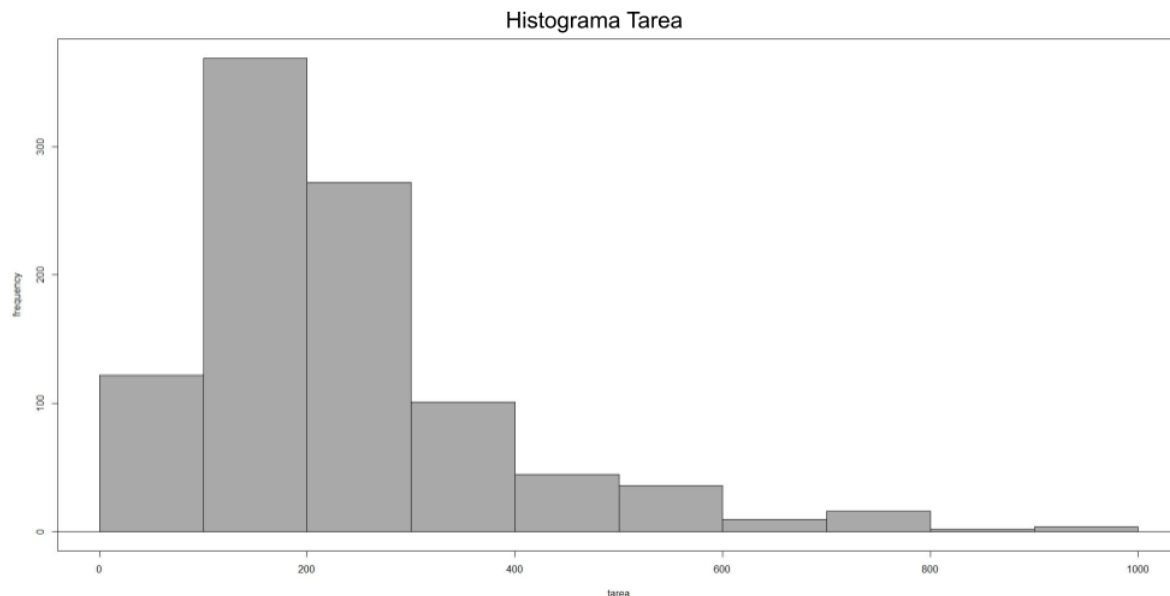


Gráfico de frequência Parea

Nos histogramas abaixo conseguimos ver que existem muitos valores entre 100 e 200.

Histograma Parea





b) Elaborar a tabela de distribuição de frequências da variável “price” (preço dos imóveis);

Class limits	f	rf	rf(%)	cf	cf(%)
[157410,591062.5)	279	0.29	28.56	279	28.56
[591062.5,1024715)	288	0.29	29.48	567	58.03
[1024715,1458368)	162	0.17	16.58	729	74.62
[1458368,1892020)	117	0.12	11.98	846	86.59
[1892020,2325673)	53	0.05	5.42	899	92.02
[2325673,2759325)	28	0.03	2.87	927	94.88
[2759325,3192978)	17	0.02	1.74	944	96.62
[3192978,3626630)	8	0.01	0.82	952	97.44
[3626630,4060283)	14	0.01	1.43	966	98.87
[4060283,4493935)	8	0.01	0.82	974	99.69
[4493935,4927588)	3	0.00	0.31	977	100.00

### Tabela de distribuição

Com a tabela fica claro que a maior frequência é entre 591062.5 e 1024715.

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

Média

```
> mean(realestateiaa$price)
[1] 1140123
```

Mediana

```
> median(realestateiaa$price)
[1] 950000
```

Moda

```
> subset(table(realestateiaa$price),
+         table(realestateiaa$price) == max(table(realestateiaa$price)))
850000
      16
```

Variância

```
> var(realestateiaa$price)
[1] 670486776668
```

Desvio Padrão

```
> sd (realestateiaa$price)
[1] 818832.6
```

Coeficiente de Variação

```
> (sd(realestateiaa$price) / mean(realestateiaa$price)) * 100
[1] 71.81967
```

## Quartis

```
> quantile(realestateiaa$price, probs = 0.25)
 25%
550000
> quantile(realestateiaa$price, probs = 0.50)
 50%
950000
> quantile(realestateiaa$price, probs = 0.75)
 75%
1485000
```

## Distancia Interquartil

```
> IQR(realestateiaa$price)
[1] 935000
```

## Percentis

```
> quantile(realestateiaa$price, c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9))
 10%    20%    30%    40%    50%    60%    70%    80%    90%
353000.0 490000.0 607056.8 799400.0 950000.0 1090000.0 1366000.0 1678800.0 2123200.0
```

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

```
One-sample z-Test

data: realestateiaa$price
z = 43.522, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 1088778 1191468
sample estimates:
mean of x
 1140123
```

O intervalo de confiança com 95% está entre 1088778 e 1191468.

e) Fazer o teste de diferença entre médias para as variáveis “parea” e “tarea”

```
Two-sample z-Test

data: realestateiaa$parea and realestateiaa$tarea
z = -14.151, p-value < 2.2e-16
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
```

A diferença entre médias para as variáveis “parea” e “tarea”, encontra-se entre 152.9867 e 235.6366.

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

```
F test to compare two variances

data: realestateiaa$parea and realestateiaa$tarea
F = 0.43855, num df = 976, denom df = 976, p-value < 2.2e-16
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
```

A diferença entre variâncias é de 0.4385501.

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

```
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
```

Atestado que as amostras são independentes

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

Teste de normalidade de Kolmogorov-Smirnov

```
Lilliefors (Kolmogorov-Smirnov) normality test  
data: price  
D = 0.14486, p-value < 0.00000000000000022
```

No teste mostra que não se deve rejeitar a normalidade da amostra.

Teste de normalidade de Anderson-Darling

Hipóteses:

H<sub>0</sub>: Os dados seguem a distribuição normal

H<sub>1</sub>: Os dados não seguem a distribuição normal

```
Anderson-Darling normality test  
data: realestateiaa$price  
A = 35.862, p-value < 0.00000000000000022
```

Como  $p\text{-value}=0.00000000000000022 < 0.05$  há evidências para rejeitarmos a hipótese H<sub>0</sub>, logo, os dados não seguem a distribuição normal.