

Primeira lista de exercícios

Ernani de Souza Cubas Neto

05/06/2022

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

Carregando as bibliotecas

```
## Carregando pacotes exigidos: car
```

```
## Carregando pacotes exigidos: carData
```

```
## Carregando pacotes exigidos: sandwich
```

```
##  
## Attaching package: 'fdth'
```

```
## The following objects are masked from 'package:stats':  
##  
## sd, var
```

```
## Carregando pacotes exigidos: lattice
```

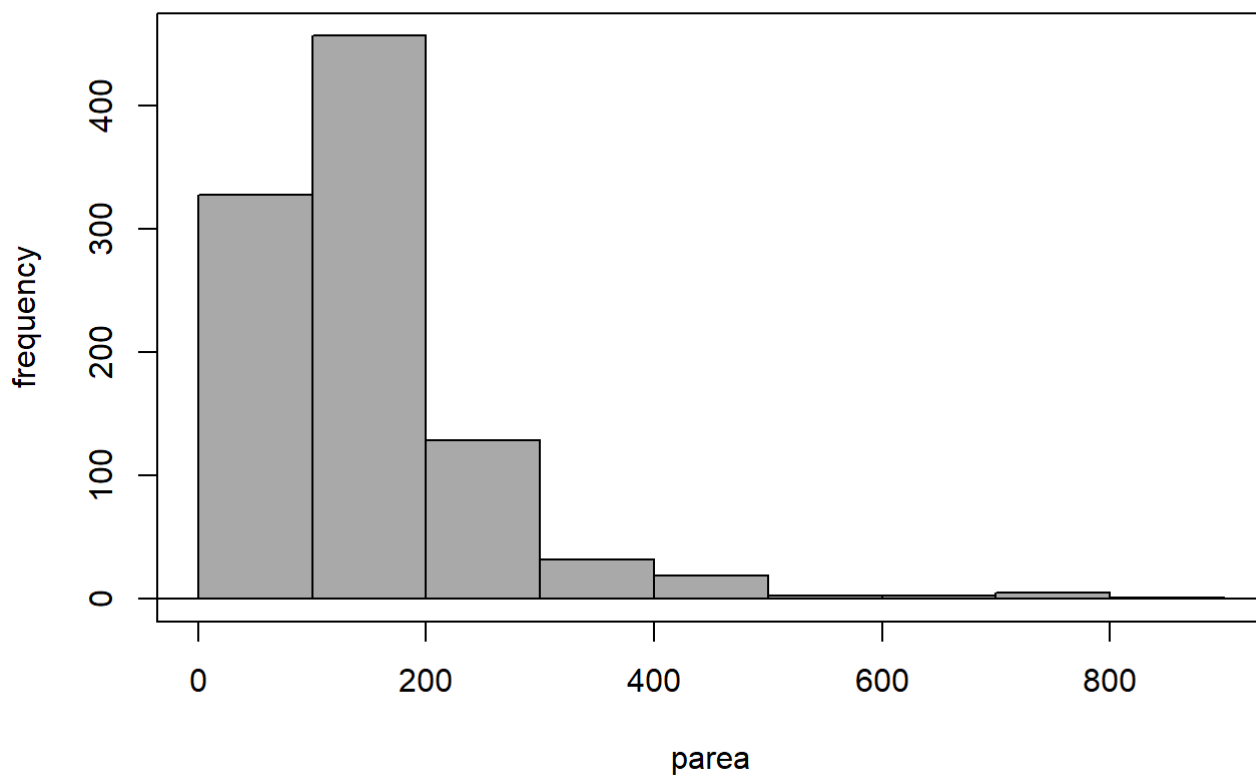
```
##  
## Attaching package: 'BSDA'
```

```
## The following objects are masked from 'package:carData':  
##  
## Vocab, Wool
```

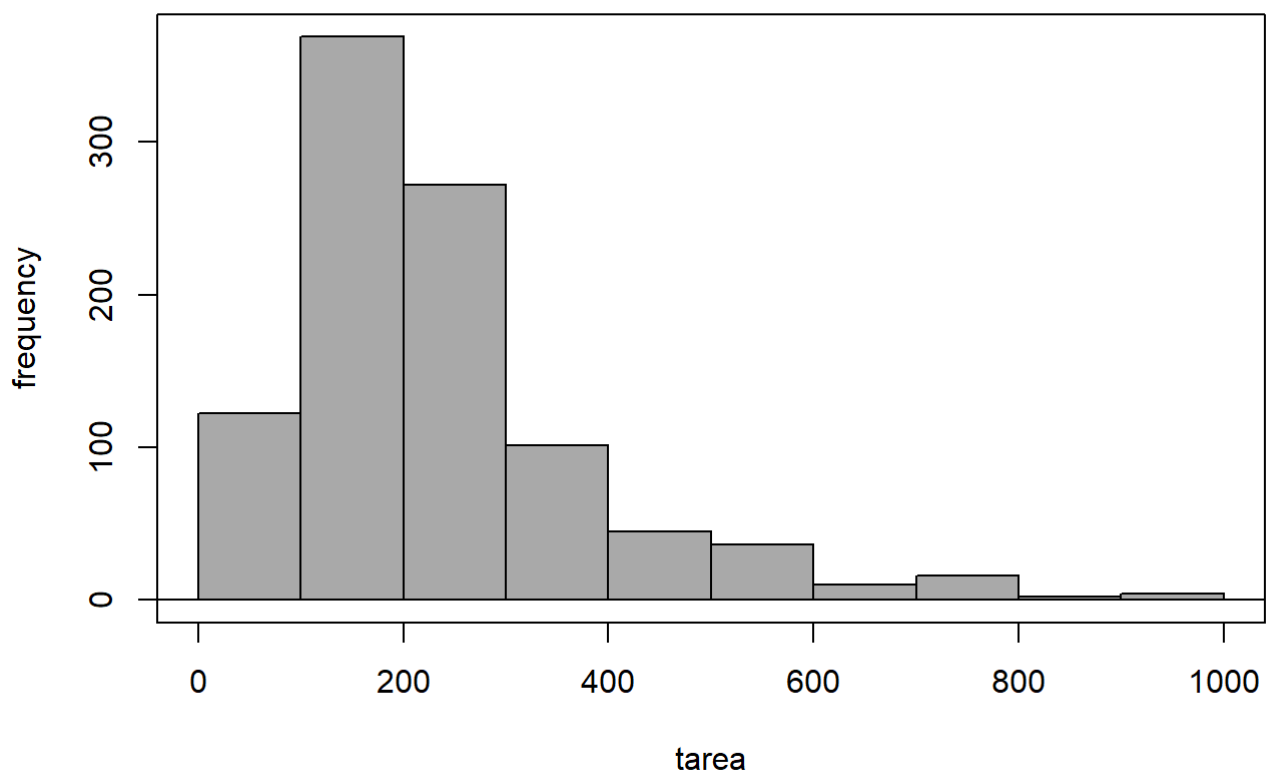
```
## The following object is masked from 'package:datasets':  
##  
## Orange
```

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

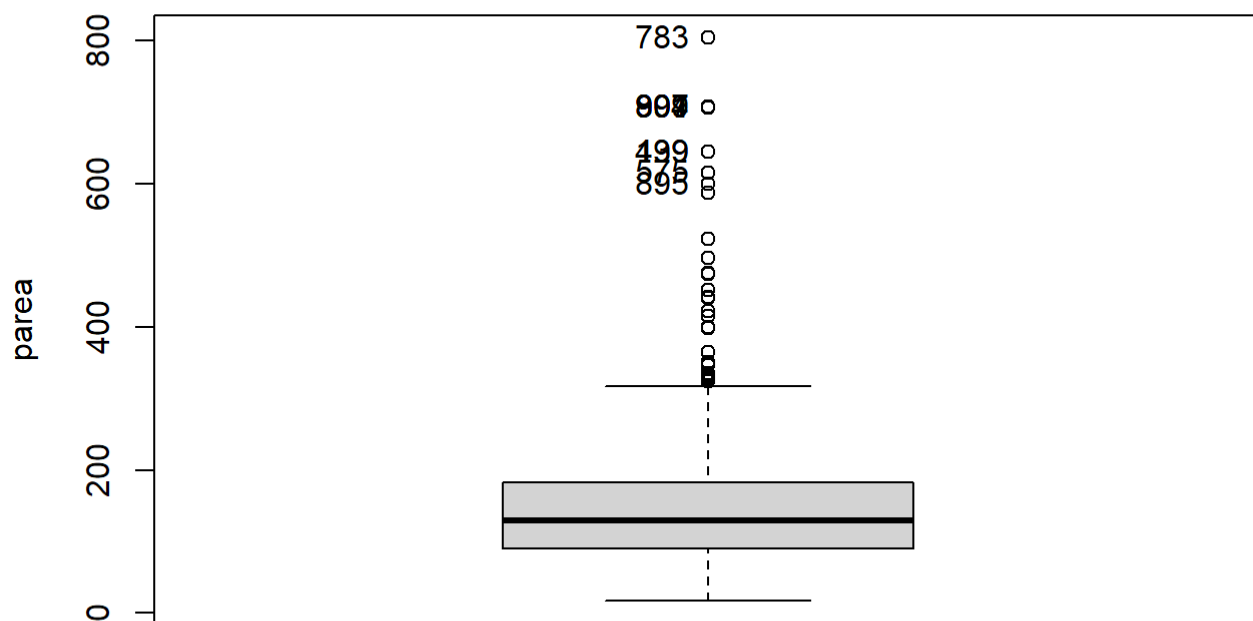
Histograma da variável “parea”



Histograma da variável “tarea”

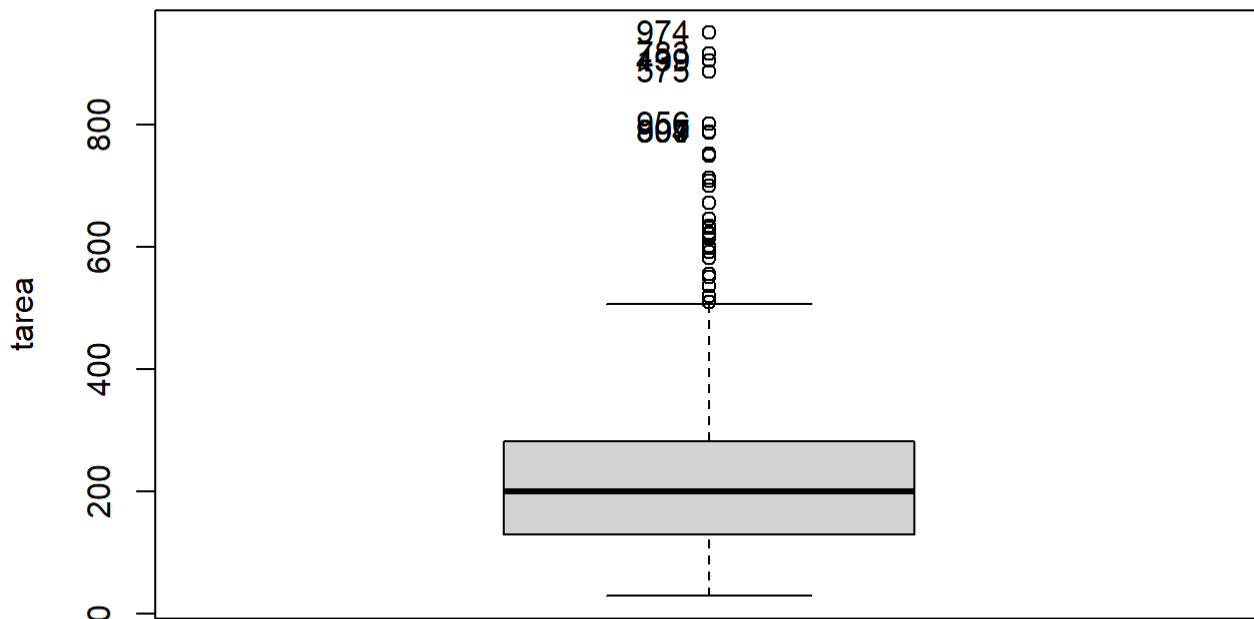


Bloxpote da variável “parea”



```
## [1] "783" "891" "904" "897" "901" "909" "139" "499" "575" "895"
```

Bloxpote da variável “tare”



```
## [1] "974" "783" "139" "499" "575" "956" "891" "904" "909" "897"
```

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

Tabela de distribuição de frequências:

##	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

Sobre a resposta: 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.

Indicador	Valor
Média	1140123.03
Mediana	950000
Moda	850000 - 16 ocorrências
Variância	670486776668.11
Desvio Padrão	818832.57
CV	71.82
Quartis	Q1- 550000 Q2- 950000 Q3- 1485000
Dist. Interq.	935000
Percentis (10%-30%)	10%- 353000 20%- 490000 30%- 607056.8
Percentis (40%-60%)	40%- 799400 50%- 950000 60%- 1090000
Percentis (70%-80%)	70%- 1366000 80%- 1678800 90%- 2123200

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 < 0.0000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1088778 1191468
## sample estimates:
## mean of x
##  1140123
```

Sobre a resposta: 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 (H_0) 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”.

```
##
## 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
```

Sobre a resposta: 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”.

```
##
## 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
```

Sobre a resposta: 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”.

```
##
## 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
```

Sobre a resposta: Como o resultado p-value é menor que 0.05, então 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
```

Sobre a resposta: Valor calculado de $D = 0.14486$ e p-value menor que 0.05, significando que não existe normalidade da amostra

Teste de normalidade de Anderson-Darling

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

Sobre a resposta: Como p-value é menor que 0.05 rejeita-se a hipótese de normalidade da variável