

Universidade Federal Rural de Pernambuco

Curso: Licenciatura em Computação – Turno: Noite

Disciplina: Estatística Exploratória I

Docente: Kleber Barros

Discente: Hugo Leonardo Luna de Castro

Atividade – Projeto para avaliação da disciplina Estatística Exploratória I

RELATÓRIO DO PROJETO - ANÁLISE ESTATÍSTICA DE FORMULÁRIO CRIADO NUMA LINGUAGEM DE PROGRAMAÇÃO

RESUMO DO PROJETO

O projeto é uma pesquisa realizada em uma empresa fictícia através de entrevista com os seus funcionários. Foi elaborado um questionário com onze perguntas simples as quais o entrevistador fará para cada funcionário. Através de suas respostas, é traçado um perfil social desses trabalhadores. Nesse projeto foi utilizando o RStudio para analisar, interpretar e produzir os gráficos dos dados obtidos. Nesta pesquisa foram analisadas variáveis qualitativas e quantitativas. As variáveis qualitativas são aquelas que não podem ser medidas. Elas se dividem em nominal e ordinal. A nominal consiste em nomes, rótulos ou categorias. Podemos citar como exemplo sexo, autodeclaração étnico-racial, etc. A ordinal pode ser agrupado em ordem. Podemos citar como exemplo a escolaridade. As variáveis quantitativas são aquelas que podem ser quantificadas, e são classificadas em discretas e contínuas. As discretas são variáveis que podem assumir somente valores inteiros. Como exemplo, temos o número de filhos, quantidades de mercadorias vendidas, etc. As contínuas são aquelas que assumem valores dentro de determinado intervalos. Podemos trazer como exemplo a altura de uma pessoa, peso, etc. Assim, a presente pesquisa buscou os seguintes dados qualitativos: sexo, autodeclaração étnico-racial, escolaridade, moradia e estado civil. Os dados quantitativos analisados foram: altura, peso, idade, número de filhos e renda.

RESULTADOS DA PESQUISA

	NOME	SEXO	ALTURA	PESO	AUTODECLARAÇÃO, ÉTNICO, RACIAL	IDADE	Nº. DE FILHOS	RENDA	ESCOLARIDADE	MORADIA	ESTADO, CIVIL
1	PAULO	MASCULINO	1.83	84	BRANCO	35	0	8500	IV	II	II
2	JOSÉ	MASCULINO	1.89	88	BRANCO	40	1	4750	III	I	III
3	LÚCIA	FEMININO	1.67	55	PARDO	34	1	4500	III	II	I
4	PAULA	FEMININO	1.70	59	PRETO	36	3	3200	II	III	I
5	ANTÔNIO	MASCULINO	1.86	86	BRANCO	28	3	3450	II	III	I
6	MARCELA	FEMININO	1.68	70	PARDO	37	2	8250	IV	II	II
7	LUCAS	MASCULINO	1.93	87	PRETO	30	1	2500	I	I	III
8	THIAGO	MASCULINO	1.84	83	BRANCO	39	3	2500	I	I	I
9	MARCOS	MASCULINO	1.80	77	PARDO	26	2	2250	I	III	III
10	CAMILA	FEMININO	1.65	52	AMARELO	23	0	2000	I	III	II
11	FERNANDA	FEMININO	1.60	56	BRANCO	30	2	3600	III	I	III
12	RAFAEL	MASCULINO	1.96	90	PRETO	37	0	3500	III	II	II
13	ANGÉLICA	FEMININO	1.70	65	PARDO	35	0	7550	IV	II	II
14	FELIPE	MASCULINO	1.70	72	BRANCO	32	2	4550	III	I	II
15	LEONARDO	MASCULINO	1.83	82	BRANCO	34	1	2300	II	III	I
16	ROBSON	MASCULINO	1.94	93	PARDO	41	0	4600	III	II	I
17	JOÃO	MASCULINO	1.78	88	PRETO	42	3	4500	III	I	I
18	OZANA	FEMININO	1.68	56	AMARELO	39	1	7600	IV	I	III
19	APOLO	MASCULINO	2.10	96	BRANCO	45	2	9650	IV	II	I
20	ROBERTO	MASCULINO	1.93	94	PARDO	27	4	3350	II	III	II

Tabela 1: Resultado da entrevista.

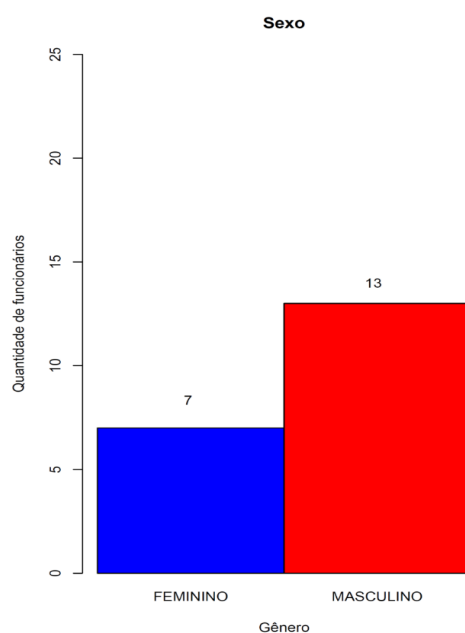
O resultado inicial da entrevista é uma tabela com as informações coletadas dos funcionários da empresa fictícia. Nessa tabela temos o nome, sexo, altura, peso, autodeclaração étnico-racial, idade, número de filhos, renda, escolaridade, moradia, estado civil. No caso da escolaridade, moradia e estado civil foram utilizadas legendas cujo significado segue abaixo:

Título	I	II	III	IV	V
Escolaridade	Nível fundamental	Nível médio	Nível superior	Pós-Graduação	Nenhum
Moradia	Alugada	Própria	Outro	-	-
Estado Civil	Casado(a)	Solteiro(a)	Outro	-	-

DADOS QUALITATIVOS

SEXO

A pesquisa revelou que o número de funcionários do sexo masculino da empresa é superior ao do sexo feminino. Os dados revelaram que na empresa 35% dos funcionários são mulheres e 65 % são homens.



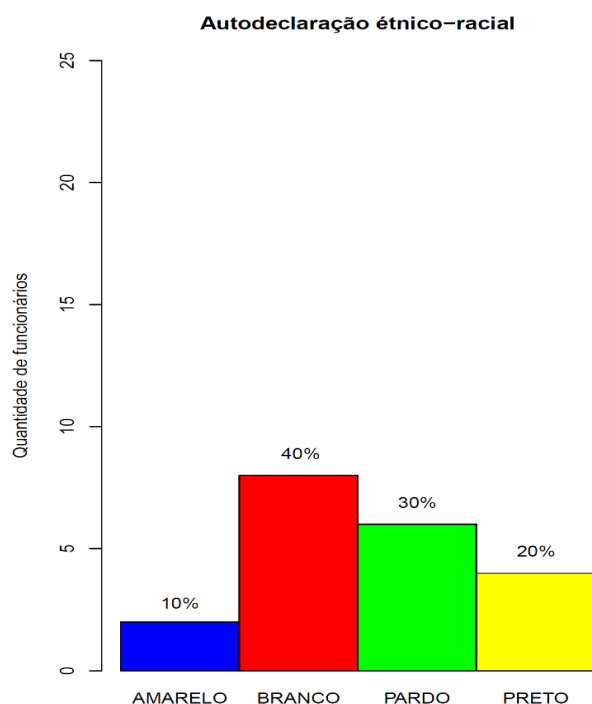
Distribuição dos funcionários por sexo.

AUTODECLARAÇÃO ÉTNICO-RACIAL

Através das respostas obtidas através do questionário, verificou-se que na empresa existem quatro grupos étnico-raciais. No total, os grupos ficaram divididos com os seguintes resultados:

	Amarelo	Branco	Pardo	Preto
Quantitativo	2	8	6	4

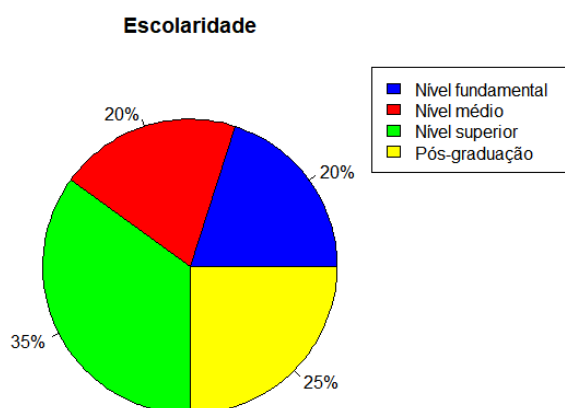
Através do gráfico de barras podemos perceber que o número de funcionários que se autodeclararam brancos é maior do que os que se declararam amarelo, pardo ou preto.



ESCOLARIDADE

O nível escolar entre os funcionários é variado, porém o predomínio está entre aqueles que possuem nível superior e pós-graduação. No total, os grupos ficaram divididos com os seguintes quantitativos de funcionários:

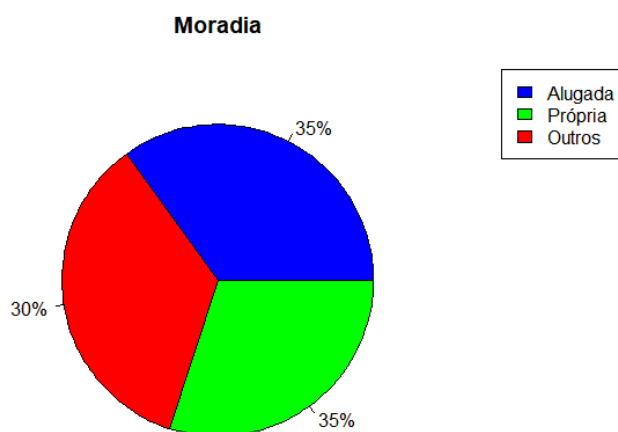
	Nível fundamental	Nível médio	Nível superior	Pós-graduação	Nenhum
Quantitativo	4	4	7	5	0



MORADIA

As informações levantadas revelaram que no quesito moradia, os funcionários estão quase no mesmo patamar. Através do gráfico de pizza podemos perceber a porcentagem dos trabalhadores que tem moradia alugada, própria e outros (Mora com os pais, amigos, etc.).

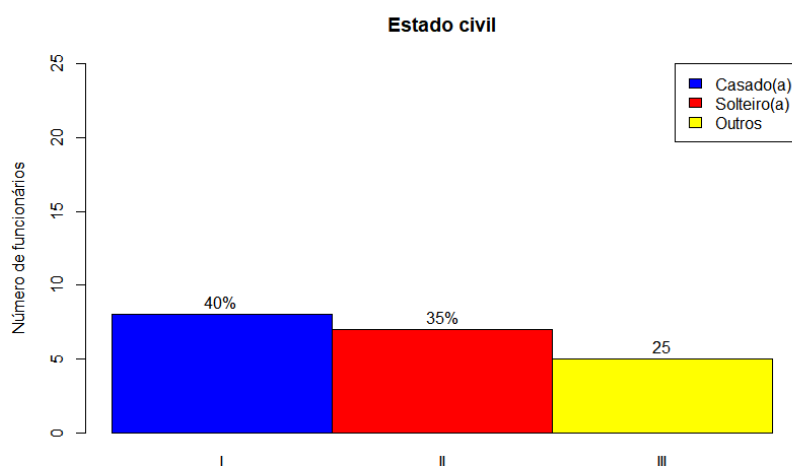
	Alugada	Própria	Outros
Quantitativo	7	7	6



ESTADO CIVIL

As informações levantadas revelaram que no quesito estado civil, os funcionários casados superam em 5% o número de funcionários solteiros. Através do gráfico de barras podemos perceber a porcentagem dos trabalhadores casados, solteiros e outros (Mora com um(a) parceiro(a), união estável, etc.).

	Casado(a)	Solteiro(a)	Outros
Quantitativo	8	7	5



DADOS QUANTITATIVOS

ALTURA

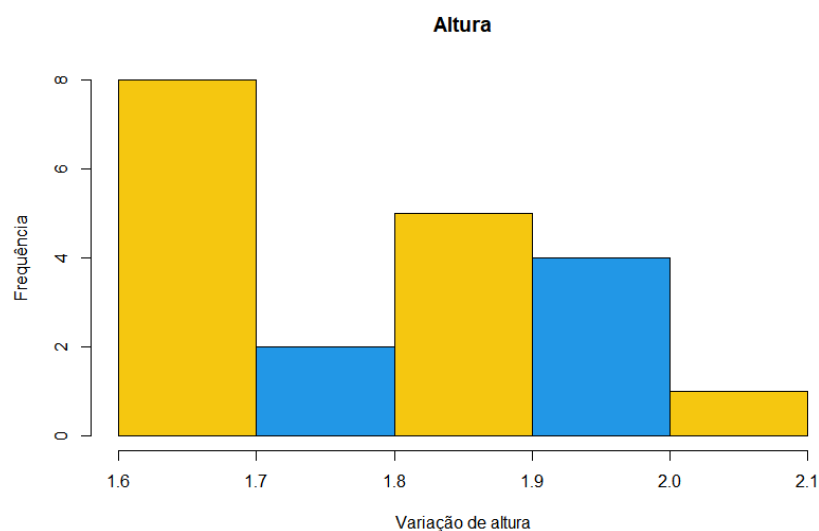
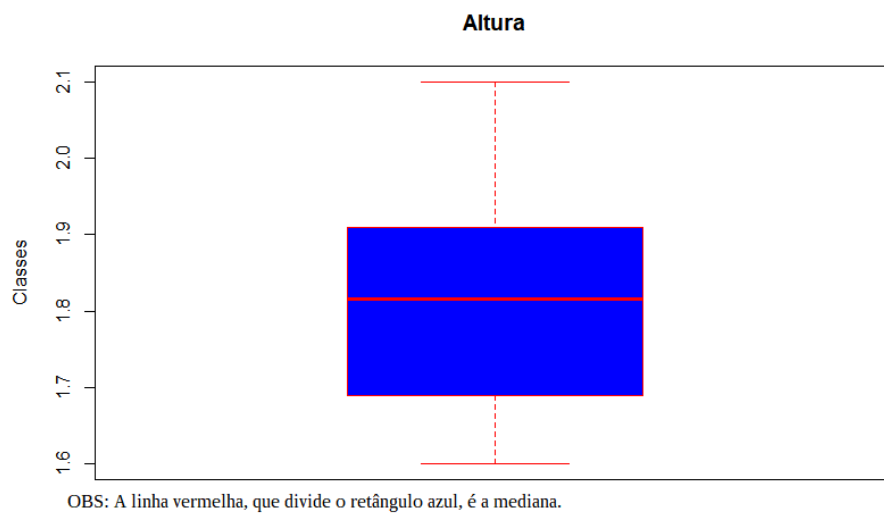
Através das respostas obtidas através do questionário, verificou-se que a altura dos funcionários da empresa é bem variada. Porém, pelo gráfico do diagrama de caixa, podemos perceber que as alturas dos funcionários estão dentro do intervalo mínimo e máximo. Já pelo gráfico do histograma podemos perceber o número de funcionários no intervalo das alturas informadas. No total, os grupos ficaram divididos nas seguintes classes:

Altura	Frequência (n_i)
1.60 ┤ 1.70	8
1.70 ┤ 1.80	2
1.80 ┤ 1.90	5
1.90 ┤ 2.00	4
2.00 ┤ 2.10	1

Resultados:

Altura mínima	Primeiro quartil	Média	Mediana	Terceiro quartil	Altura máxima
1.60	1.695	1.80	1.85	1.900	2.10

Variância	Desvio padrão
0.01697132	0.130274



PESO

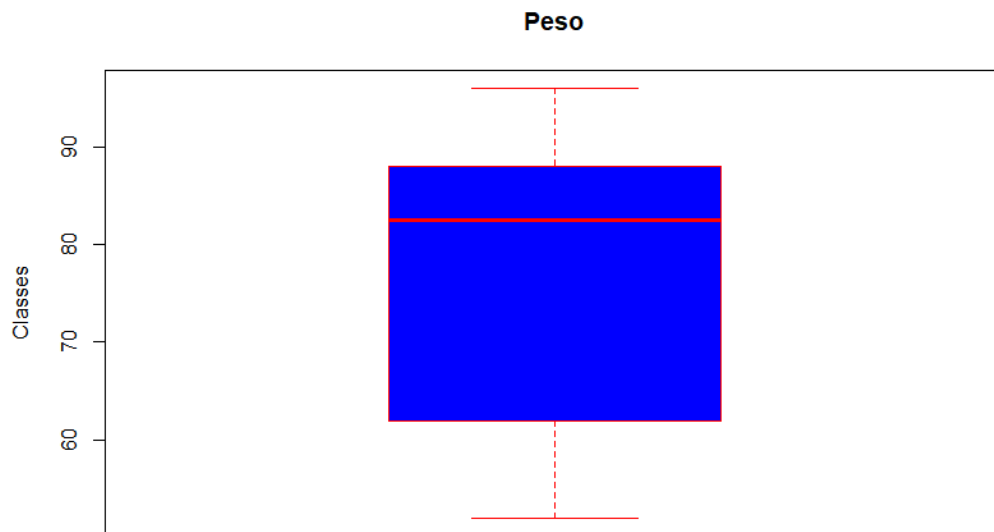
Pelas informações levantadas podemos fazer uma análise do peso dos funcionários da empresa analisada. O gráfico de diagrama de caixa nos mostra que o peso dos funcionários estão dentro do intervalo mínimo e máximo. Desse modo, no quesito peso, os funcionários apresentaram os seguintes dados:

Peso	Frequência (n_i)
50 60	5
60 70	2
70 80	2
80 90	8
90 100	3

Resultados:

Peso mínima	Primeiro quartil	Média	Mediana	Terceiro quartil	Peso máxima
52	63.5	76.65	82.50	88	96

Variância	Desvio padrão
217.8184	14.75867



OBS: A linha vermelha, que divide o retângulo azul, é a mediana.

IDADE

As informações levantadas revelaram que no quesito idade, os funcionários apresentaram os seguintes dados:

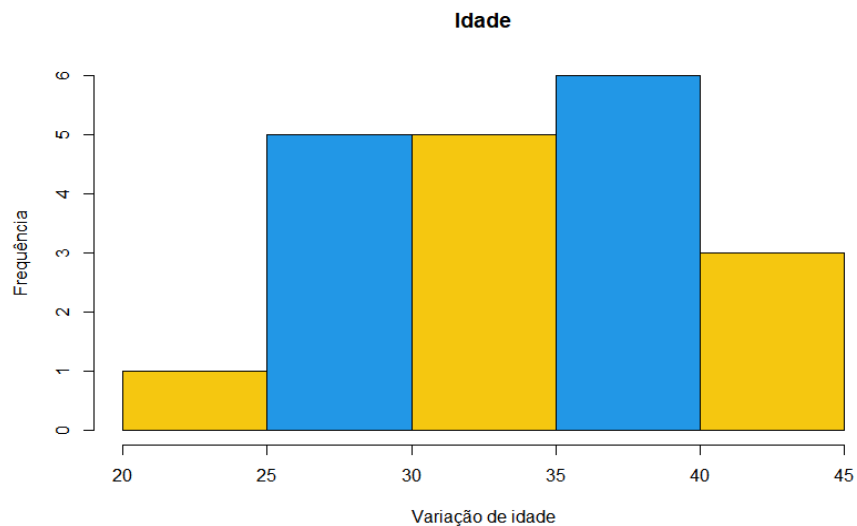
Idade	Frequência (n_i)
20 25	1
25 30	5
30 35	5
35 40	6
40 45	3

Resultados:

Idade mínima	Primeiro quartil	Média	Mediana	Terceiro quartil	Idade máxima
23	30	34.5	35	39	45

Variância	Desvio padrão
33.94737	5.826437

Pelo gráfico do histograma podemos perceber o número de funcionários no eixo y (frequência) e no eixo x (variação de idade) os intervalos de idade desses trabalhadores.



NÚMERO DE FILHOS

Através das respostas obtidas através do questionário, verificou-se que na empresa o número de funcionários que tem filhos é bem superior

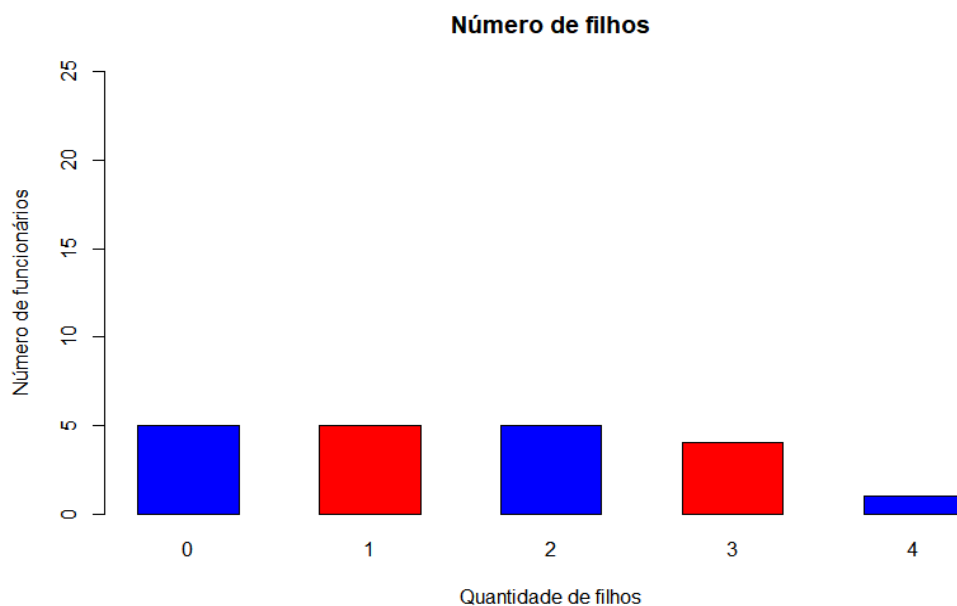
ao número de funcionários que tem filhos. O percentual de trabalhadores, dessa empresa, que tem filhos foi de 75% e dos que não tem filhos foi de 25%. Porém, o número de filhos oscilou muito entre 0, 1 e 2. Esse dado revela uma realidade no país em que o crescimento vegetativo, segundo estudos, embora ainda positivo, é lento e estável. No total, os grupos ficaram divididos no seguinte rol:

Número de filhos	0	1	2	3	4
Número de funcionários	5	5	5	4	1

Resultados:

Mínimo	Primeiro quartil	Média	Mediana	Terceiro quartil	Máximo
0	0.75	1.55	1.50	2.25	4

Variância	Desvio padrão
1.523684	1.234376



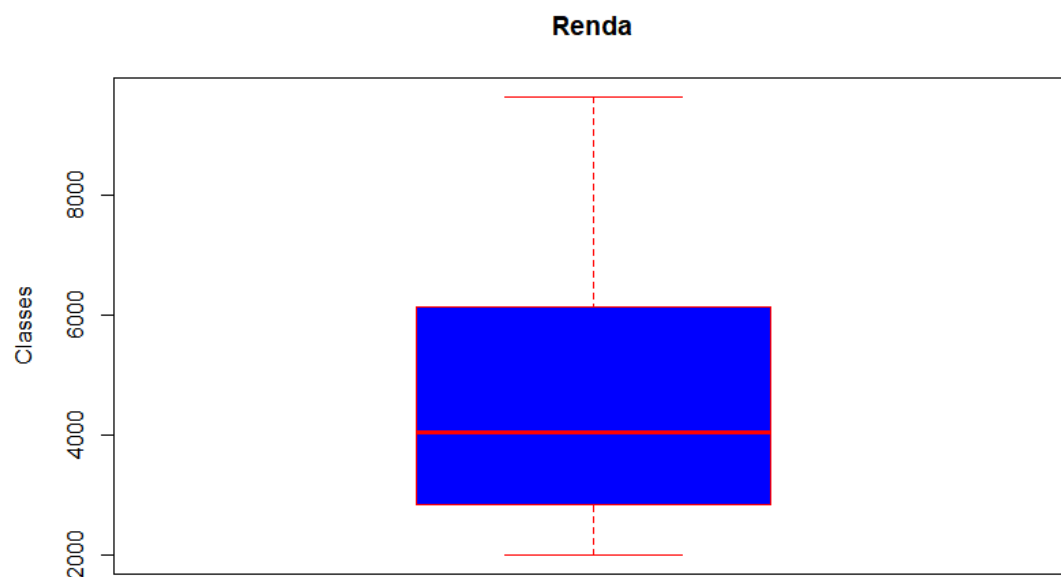
RENDA

As informações levantadas revelaram que no quesito renda nos mostra que a renda dos funcionários apresentam uma variação de aproximadamente 21% entre menor e a maior renda informadas. Sobre a renda dos funcionários, obteve-se os seguintes dados:

Renda mínima	Primeiro quartil	Média	Mediana	Terceiro quartil	Renda máxima
2000	3025	4655	4050	5450	9650

Variância	Desvio padrão
5535237	2352.708

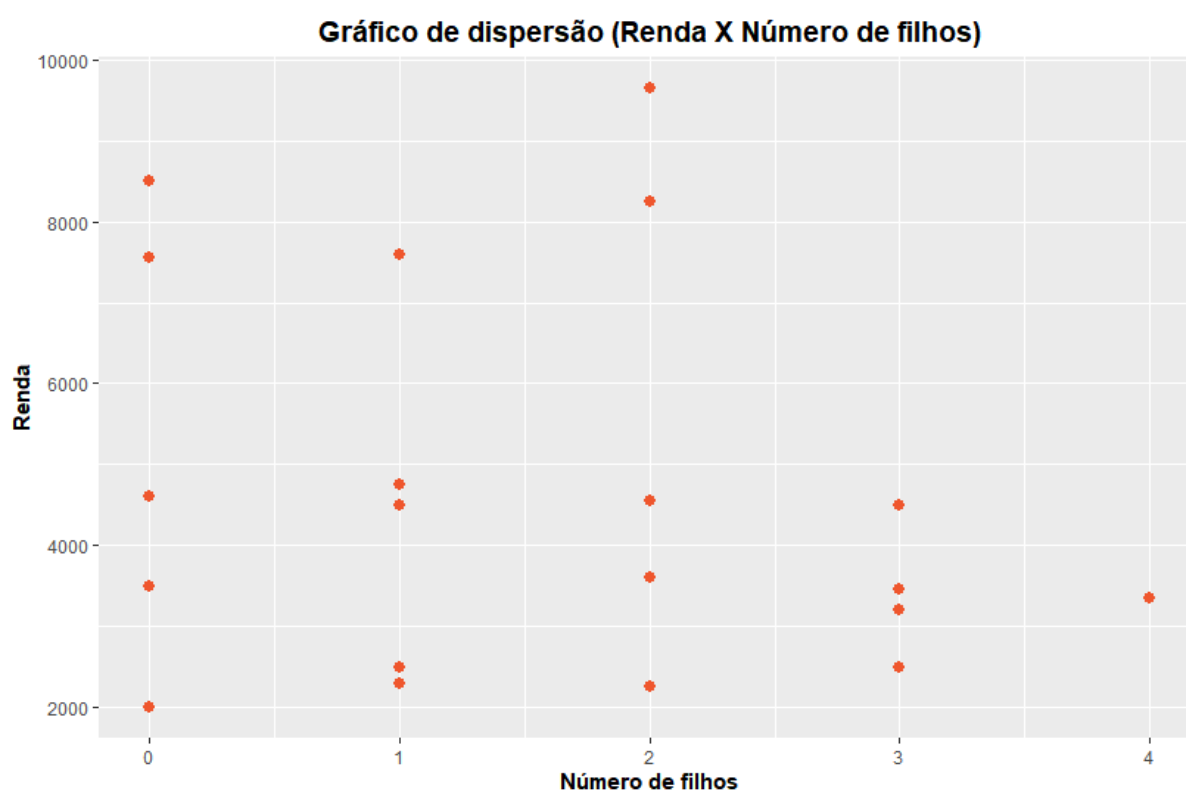
O gráfico de diagrama de caixa nos mostra que a renda dos funcionários estão dentro do intervalo mínimo e máximo.



OBS: A linha vermelha, que divide o retângulo azul, é a mediana.

RENDAX NÚMERO DE FILHOS

Através do cruzamento dos dados de renda e número de filhos, podemos fazer uma análise dessas informações para saber se existe uma relação entre a quantidade de filhos e a renda dos funcionários. Assim, foi elaborado um gráfico de dispersão para averiguar a relação que existe entre esses dados.



Segundo o gráfico, podemos perceber que os funcionários com as melhores rendas têm entre 0 e 2 filhos. Por outro lado, os funcionários com renda mais baixa têm entre 3 e 4 filhos.

APÊNDICE

Código elaborado no Python para coleta de dados e produção de arquivo de texto para o Rstudio:

with open ('pesquisa.txt', 'w') as arquivo:

```
listnome = []
```

```
listasexo = []
```

```
listaltura = []
```

```
listpeso = []
```

```
listcordapele = []
```

```
listidade= []
```

```
listfilhos = []
```

```
listrenda = []
```

```
listescolaridade = []
```

```
listresidencia = []
```

```
listec = []
```

```
for i in range (1, 21, 1):
```

```
    nome = input('Digite o nome do funcionário: ').upper()
```

```
    listnome.append(nome)
```

```
    sexo = input('Sexo: ').upper()
```

```
    listasexo.append(sexo)
```

```
    altura = float(input('Digite a altura: '))
```

```
    listaltura.append(altura)
```

```
    peso = float(input('Digite o seu peso: '))
```

```
    listpeso.append(peso)
```

```
    cordapele = input('Digite a cor da sua pele: ').upper()
```

```
    listcordapele.append(cordapele)
```

```
    idade = int(input('Digite a sua idade: '))
```

```
    listidade.append(idade)
```

```
    filhos = int(input('Quantos filhos ele(a) tem: '))
```

```
    listfilhos.append(filhos)
```

```
    renda = float(input('Digite a renda: '))
```

```

    listrenda.append(renda)

    escolaridade = str(input('Qual o nível de escolaridade: \n[I] Nível Fundamental \n[II]
Nível Médio \n[III] Nível Superior \n[IV] Pós-Graduação \n[V] Nenhum \nResposta:
')).upper()

    listescolaridade.append(escolaridade)

    residencia = str(input('Moradia: \n[I] Alugada \n[II] Própria \n[III] Outro \nResposta:
')).upper()

    listresidencia.append(residencia)

    ec = str(input('Estado civil: \n[I] Casado(a) \n[II] Solteiro(a) \n[III] Outros \nResposta:
')).upper()

    listec.append(ec)

    print('NOME, SEXO, ALTURA, PESO, AUTODECLARAÇÃO ÉTNICO-RACIAL,
IDADE, Nº DE FILHOS, RENDA, ESCOLARIDADE, MORADIA, ESTADO CIVIL',
file=arquivo)

    print(f'{listnome[0]},{listasexo[0]},{listaltura[0]},{listpeso[0]},{listcordapele[0]},
{listidade[0]},{listfilhos[0]},{listrenda[0]},{listescolaridade[0]},{listresidencia[0]},
{listec[0]}', file=arquivo)

    print(f'{listnome[1]},{listasexo[1]},{listaltura[1]},{listpeso[1]},{listcordapele[1]},
{listidade[1]},{listfilhos[1]},{listrenda[1]},{listescolaridade[1]},{listresidencia[1]},
{listec[1]}', file=arquivo)

    print(f'{listnome[2]},{listasexo[2]},{listaltura[2]},{listpeso[2]},{listcordapele[2]},
{listidade[2]},{listfilhos[2]},{listrenda[2]},{listescolaridade[2]},{listresidencia[2]},
{listec[2]}', file=arquivo)

    print(f'{listnome[3]},{listasexo[3]},{listaltura[3]},{listpeso[3]},{listcordapele[3]},
{listidade[3]},{listfilhos[3]},{listrenda[3]},{listescolaridade[3]},{listresidencia[3]},
{listec[3]}', file=arquivo)

    print(f'{listnome[4]},{listasexo[4]},{listaltura[4]},{listpeso[4]},{listcordapele[4]},
{listidade[4]},{listfilhos[4]},{listrenda[4]},{listescolaridade[4]},{listresidencia[4]},
{listec[4]}', file=arquivo)

    print(f'{listnome[5]},{listasexo[5]},{listaltura[5]},{listpeso[5]},{listcordapele[5]},
{listidade[5]},{listfilhos[5]},{listrenda[5]},{listescolaridade[5]},{listresidencia[5]},
{listec[5]}', file=arquivo)

```

```

    print(f'{listnome[6]},{listasexo[6]},{listaltura[6]},{listpeso[6]},{listcordapele[6]},
{listidade[6]},{listfilhos[6]},{listrenda[6]},{listescolaridade[6]},{listresidencia[6]},
{listec[6]}', file=arquivo)

    print(f'{listnome[7]},{listasexo[7]},{listaltura[7]},{listpeso[7]},{listcordapele[7]},
{listidade[7]},{listfilhos[7]},{listrenda[7]},{listescolaridade[7]},{listresidencia[7]},
{listec[7]}', file=arquivo)

    print(f'{listnome[8]},{listasexo[8]},{listaltura[8]},{listpeso[8]},{listcordapele[8]},
{listidade[8]},{listfilhos[8]},{listrenda[8]},{listescolaridade[8]},{listresidencia[8]},
{listec[8]}', file=arquivo)

    print(f'{listnome[9]},{listasexo[9]},{listaltura[9]},{listpeso[9]},{listcordapele[9]},
{listidade[9]},{listfilhos[9]},{listrenda[9]},{listescolaridade[9]},{listresidencia[9]},
{listec[9]}', file=arquivo)

    print(f'{listnome[10]},{listasexo[10]},{listaltura[10]},{listpeso[10]},{listcordapele[10]},
{listidade[10]},{listfilhos[10]},{listrenda[10]},{listescolaridade[10]},{listresidencia[10]},
{listec[10]}', file=arquivo)

    print(f'{listnome[11]},{listasexo[11]},{listaltura[11]},{listpeso[11]},{listcordapele[11]},
{listidade[11]},{listfilhos[11]},{listrenda[11]},{listescolaridade[11]},{listresidencia[11]},
{listec[11]}', file=arquivo)

    print(f'{listnome[12]},{listasexo[12]},{listaltura[12]},{listpeso[12]},{listcordapele[12]},
{listidade[12]},{listfilhos[12]},{listrenda[12]},{listescolaridade[12]},{listresidencia[12]},
{listec[12]}', file=arquivo)

    print(f'{listnome[13]},{listasexo[13]},{listaltura[13]},{listpeso[13]},{listcordapele[13]},
{listidade[13]},{listfilhos[13]},{listrenda[13]},{listescolaridade[13]},{listresidencia[13]},
{listec[13]}', file=arquivo)

    print(f'{listnome[14]},{listasexo[14]},{listaltura[14]},{listpeso[14]},{listcordapele[14]},
{listidade[14]},{listfilhos[14]},{listrenda[14]},{listescolaridade[14]},{listresidencia[14]},
{listec[14]}', file=arquivo)

    print(f'{listnome[15]},{listasexo[15]},{listaltura[15]},{listpeso[15]},{listcordapele[15]},
{listidade[15]},{listfilhos[15]},{listrenda[15]},{listescolaridade[15]},{listresidencia[15]},
{listec[15]}', file=arquivo)

    print(f'{listnome[16]},{listasexo[16]},{listaltura[16]},{listpeso[16]},{listcordapele[16]},
{listidade[16]},{listfilhos[16]},{listrenda[16]},{listescolaridade[16]},{listresidencia[16]},
{listec[16]}', file=arquivo)

```

```

print(f'{listnome[17]},{listasexo[17]},{listaltura[17]},{listpeso[17]},{listcordapele[17]},
{listidade[17]},{listfilhos[17]},{listrenda[17]},{listescolaridade[17]},{listresidencia[17]},
{listec[17]}' , file=arquivo)

print(f'{listnome[18]},{listasexo[18]},{listaltura[18]},{listpeso[18]},{listcordapele[18]},
{listidade[18]},{listfilhos[18]},{listrenda[18]},{listescolaridade[18]},{listresidencia[18]},
{listec[18]}' , file=arquivo)

print(f'{listnome[19]},{listasexo[19]},{listaltura[19]},{listpeso[19]},{listcordapele[19]},
{listidade[19]},{listfilhos[19]},{listrenda[19]},{listescolaridade[19]},{listresidencia[19]},
{listec[19]}' , file=arquivo)

```

Código do Rstudio para análise dos dados e elaboração dos gráficos:

```

setwd('C:\\Users\\HUGO CASTRO\\Desktop\\2021.1\\Estatística\\Projeto full')

df = read.csv("pesquisa.txt", header = T)
prop.table(table(df$SEXO))
prop.table(table(df$AUTODECLARAÇÃO.ÉTNICO.RACIAL))
prop.table(table(df$ESCOLARIDADE))
prop.table(table(df$MORADIA))
prop.table(table(df$ESTADO.CIVIL))
summary(df$ALTURA)
summary(df$PESO)
summary(df$IDADE)
summary(df$Nº.DE.FILHOS)
summary(df$RENDIA)
var(df$ALTURA)
var(df$PESO)
var(df$IDADE)
var(df$Nº.DE.FILHOS)
var(df$RENDIA)
sd(df$ALTURA)
sd(df$PESO)
sd(df$IDADE)
sd(df$Nº.DE.FILHOS)
sd(df$RENDIA)
barplot(table(df$SEXO), col=c("blue","red"),
        ylim=c(0,25),
        space=0, width=c(.1,.1),
        main="Sexo",
        xlab="Gênero", ylab="Quantidade de funcionários")
text(locator(n=2),c("7","13"))
barplot(table(df$AUTODECLARAÇÃO.ÉTNICO.RACIAL),
        col=c("blue","red","green","yellow"),

```



```

ylim=c(0,25),
space=0, width=c(.1,.1),
main="Autodeclaração étnico-racial",
xlab=" ", ylab="Quantidade de funcionários")
text(locator(n=4),c("10%","40%","30%","20%"))
pie(table(df$ESCOLARIDADE), main="Escolaridade",
labels=c("20%","20%","35%","25%"),col=c("Blue","red","green","yellow"))
legend("topright",fill=c("blue","red","green","yellow"),legend=c("Nível fundamental","Nível
médio","Nível superior","Pós-graduação"))
pie(table(df$MORADIA), main="Moradia", labels=c("35%","30%","35%"),
col=c("Blue","red","green"))
legend("topright",fill=c("blue","green","red"),legend=c("Alugada","Própria","Outros"))
barplot(table(df$ESTADO.CIVIL), col=c("blue","red","yellow"),
ylim=c(0,25),
space=.0, width=c(.2,.2),
main="Estado civil",
xlab=" ", ylab="Número de funcionários")
text(locator(n=3),c("40%","35%","25%"))
legend("topright",fill=c("blue","red","yellow"),legend=c("Casado(a)","Solteiro(a)","Outros"))
boxplot(df$ALTURA, col="blue", main="Altura", ylab="Classes", border="red")
hist(df$ALTURA, main="Altura", col=c(7,4),
xlab="Variação de altura", ylab="Frequência")
boxplot(df$PESO, col="blue", main="Peso", ylab="Classes", border="red")
hist(df$IDADE, main="Idade", col=c(7,4),
xlab="Variação de idade", ylab="Frequência")
barplot(table(df$Nº.DE.FILHOS), col=c("blue","red"),
ylim=c(0,25),
space=.8, width=c(.2,.2),
main="Número de filhos",
xlab="Quantidade de filhos", ylab="Número de funcionários")
text(locator(n=5),c("5","5","5","4","1"))
boxplot(df$RENDIA, col="blue", main="Renda", ylab="Classes", border="red")
library(ggplot2)
library(esquisse)
library(ggplot2)
ggplot(pesquisa) +
aes(x = Nº.DE.FILHOS, y = RENDIA) +
geom_point(shape = "circle", size = 2.45, colour = "#EF562D") +
labs(x = "Número de filhos", y = "Renda", title = "Gráfico de dispersão (Renda X Número de
filhos)") +
theme_gray() +
theme(plot.title = element_text(size = 14L, face = "bold", hjust = 0.5), axis.title.y =
element_text(face = "bold"),
axis.title.x = element_text(face = "bold"))

```