

Atividade avaliativa 1

Questão 1:

a -

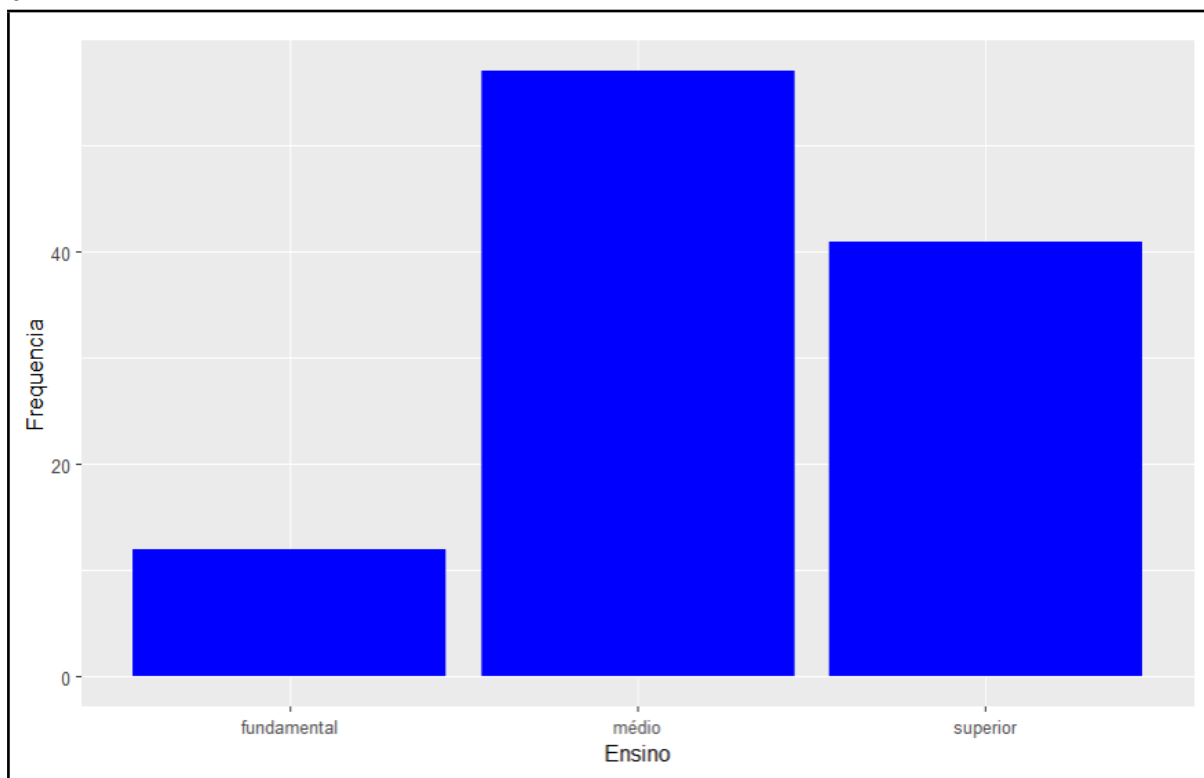
```
não sim
70 40

> freq.tabela <- table(seguro$sinistro) # frequencia absoluta
> freq.tabela

não sim
70 40
>
> relativa <- prop.table(freq.tabela) # frequencia relativa
> relativa

      não      sim
0.6363636 0.3636364
> |
```

b -



c -

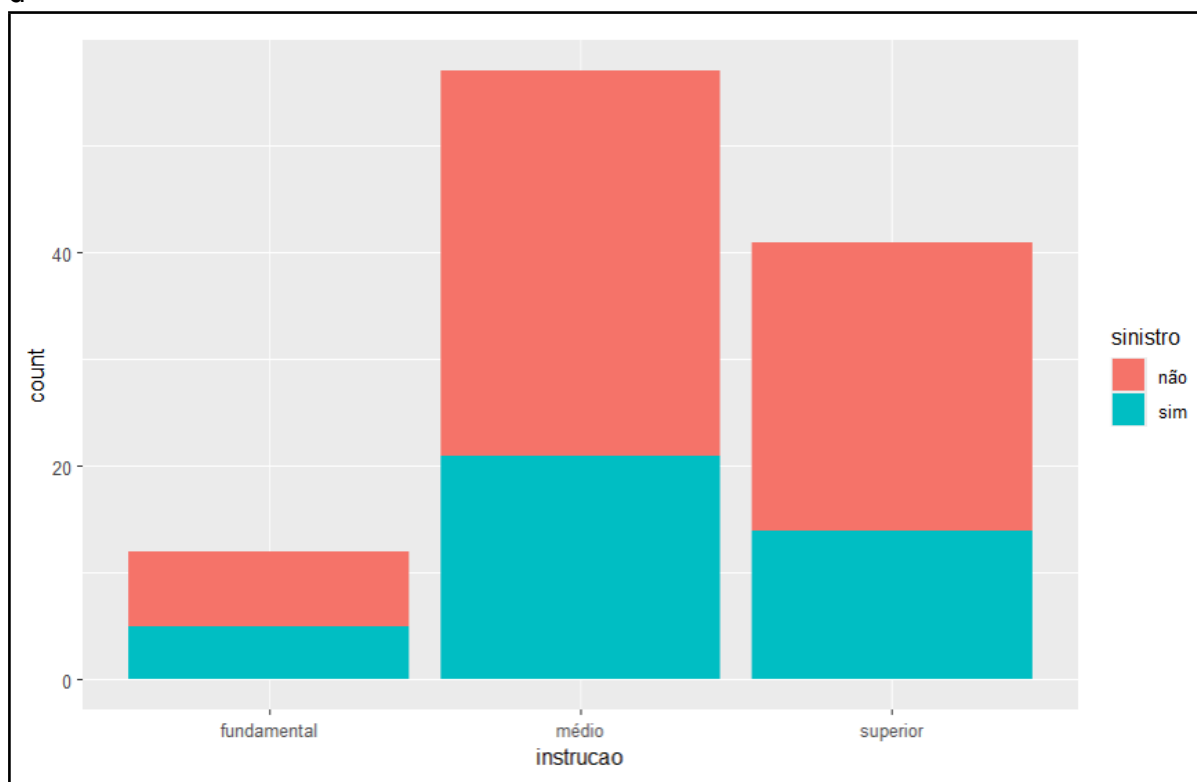
```
      não sim  
casado   37  21  
solteiro 33  19
```

```
> cruzada <- table(seguro$estado, seguro$sinistro)  
> cruzada
```

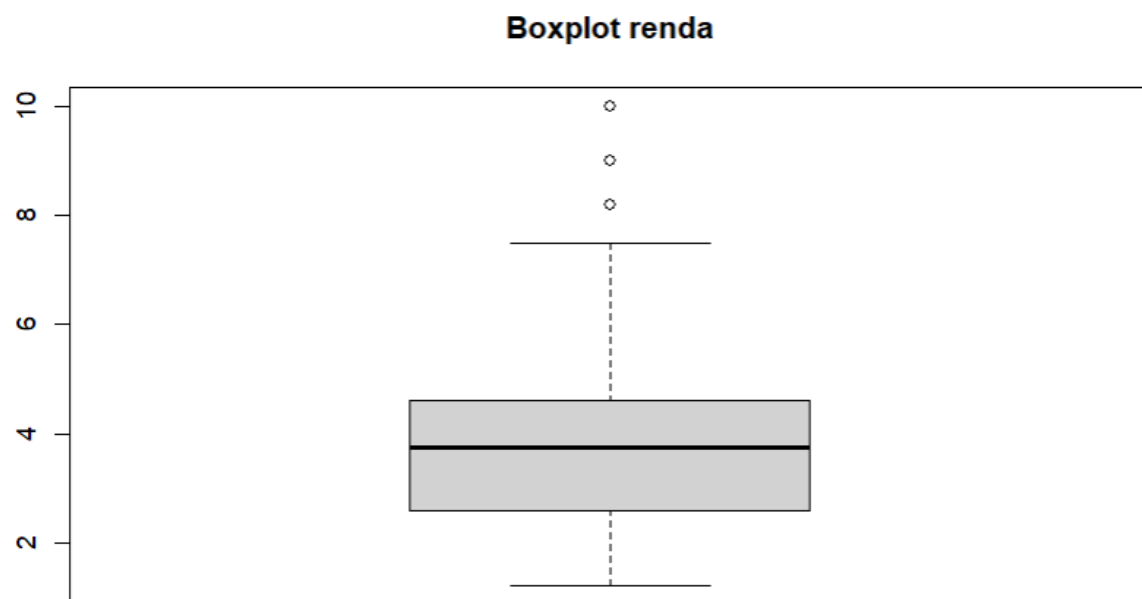
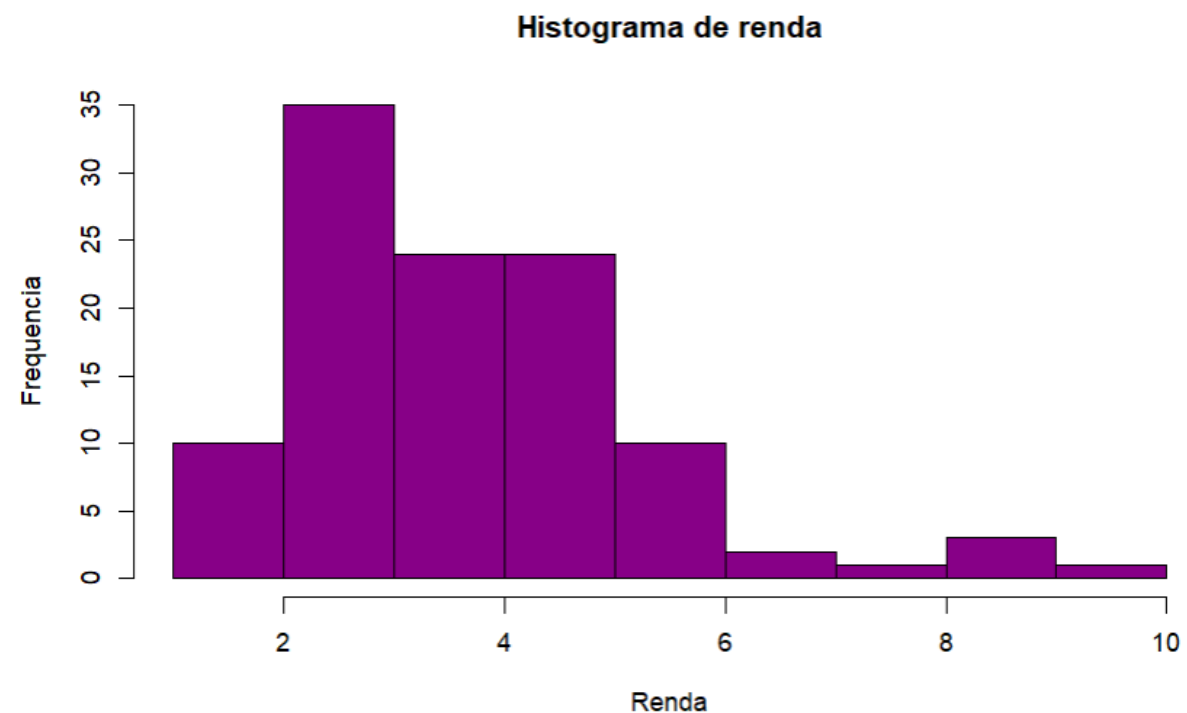
```
      não sim  
casado   37  21  
solteiro 33  19
```

```
>
```

d -



e -



```
> iqr <- IQR(seguro$renda)
> q1 <- quantile(seguro$renda, 0.25)
> q3 <- quantile(seguro$renda, 0.75)
> inf <- q1 - 1.5 * iqr
> sup <- q3 + 1.5 * iqr
> outliers <- seguro$renda[seguro$renda < inf | seguro$renda > sup]
> outliers
[1] 9.0 8.2 8.2 10.0
> |
```

Há valores discrepantes.

```
> library("e1071")
> skewness(seguro$renda)
[1] 1.137541
> |
```

Há assimetria positiva.

f -

```
> renda <- seguro$renda
> mean(renda)      # média
[1] 3.816364
> median(renda)    # mediana
[1] 3.75
> sd(renda)        # desvio padrão
[1] 1.624781
> coefvari <- (sd(renda)/mean(renda))*100    # coeficiente variacao
> quantile(seguro$renda, probs=c(0.25,0.75)) # quartis
      25%    75%
2.600 4.575
> |
```

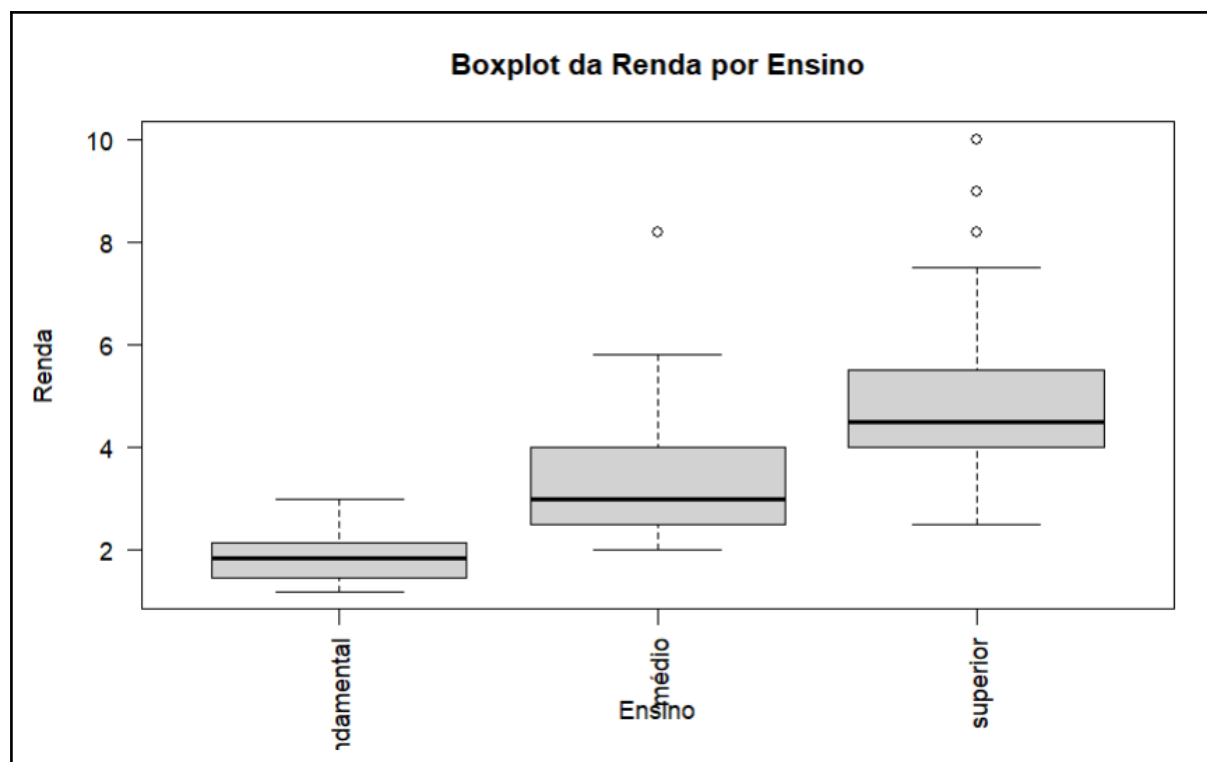
g -

```
> renda <- seguro$renda
> faixas <- cut(
+   renda,
+   breaks = c(0, 1.9, 3.9, 6.9, Inf),
+   labels = c("Até 1.9", "2.0 até 3.9", "4.0 até 6.9", "7.0 ou mais")
+ )
> tabelafaixas <- table(faixas)
> frequencia <- prop.table(tabelafaixas) * 100
> tabela <- data.frame(
+   faixas = names(tabelafaixas),
+   frequencias = as.vector(tabelafaixas),
+   frqrelativa = round(as.vector(frequencia), 2)
+ )
> tabela
```

	faixas	frequencias	frqrelativa
1	Até 1.9	7	6.36
2	2.0 até 3.9	53	48.18
3	4.0 até 6.9	45	40.91
4	7.0 ou mais	5	4.55

```
> |
```

h -



i -

```
[1] "fundamental"
[1] "media: 1.875"
[1] "mediana: 1.85"
[1] "q1: 1.525"
[1] "q3: 2.125"
[1] "desvio padrao: 0.51720402163943"
[1] "coeficiente de variacao (%): 27.5842144874363"
[1] "-----"
[1] "médio"
[1] "media: 3.43157894736842"
[1] "mediana: 3"
[1] "q1: 2.5"
[1] "q3: 4"
[1] "desvio padrao: 1.16437565960979"
[1] "coeficiente de variacao (%): 33.9311925346411"
[1] "-----"
[1] "superior"
[1] "media: 4.91951219512195"
[1] "mediana: 4.5"
[1] "q1: 4"
[1] "q3: 5.5"
[1] "desvio padrao: 1.62268596964957"
[1] "coeficiente de variacao (%): 32.9846924916372"
[1] "-----"
> |
```

Questão 2:

```
> # Considerando N = 110
> N = 110
> # z = 1,96; d = 6 e o = 14
> z = 1.96
> d = 6
> o = 14
> n0 = ((z^2)*(o^2))/(d^2)
> n0
[1] 20.91538
> n = n0/(1 + n0/N)
> n
[1] 17.57388
> |
```

Ou seja, o tamanho da amostra para estimar a idade média é de 18 pessoas.

Questão 3:

Média amostral:

```
> idades = seguro$idade
> idadessorteadas <- sample(idades, size=18, replace=FALSE)
> idadessorteadas
[1] 42 30 62 22 57 37 54 21 60 28 40 58 29 48 33 19 18 30
> erro <- 6
> media <- mean(idadessorteadas)
> inf <- media - erro
> sup <- media + erro
> print(paste("media:", media))
[1] "media: 38.2222222222222"
> print(paste("erro inferior:", inf))
[1] "erro inferior: 32.2222222222222"
> print(paste("erro superior:", sup))
[1] "erro superior: 44.2222222222222"
>
```

Média total:

```
> idades = seguro$idade
> erro <- 6
> media <- mean(idades)
> inf <- media - erro
> sup <- media + erro
> print(paste("media:", media))
[1] "media: 37.7363636363636"
> print(paste("erro inferior:", inf))
[1] "erro inferior: 31.7363636363636"
> print(paste("erro superior:", sup))
[1] "erro superior: 43.7363636363636"
> |
```

A média populacional está dentro da média amostral

Apendice:

Achei essa tabela, ajudou bastante.

Medidas de tendência central	Descrição
<u>mean()</u>	Média aritmética
<u>median()</u>	Mediana
Medidas de dispersão	
<u>sd()</u>	Desvio-padrão
<u>range()</u>	intervalo
<u>IQR()</u>	Intervalo interquantil
<u>var()</u>	Variância
<u>cov()</u>	Covariância
Medidas de posição	
<u>quantile()</u>	Quartis
Medidas de Forma	
<u>Kurtosis()</u>	kurtosis
<u>skewness()</u>	Assimetria
Outros medidas	
<u>max()</u>	Máximo
<u>cor()</u>	Correlação
<u>min()</u>	Mínimo

Código utilizado:

```
# Questão 1
# Questão 1
# Questão 1

# letra a

freq.tabela <- table(seguro$sinistro) # frequencia absoluta
freq.tabela

relativa <- prop.table(freq.tabela) # frequencia relativa
relativa

# letra b
install.packages("ggplot2")
library(ggplot2)
freq_instrucao <- as.data.frame(table(seguro$instrucao))
colnames(freq_instrucao) <- c("Ensino", "Frequencia")

ggplot(freq_instrucao, aes(x = Ensino, y = Frequencia)) +
  geom_bar(stat = "identity", fill = "blue")

# letra c
```



```

cruzada <- table(seguro$estado, seguro$sinistro)
cruzada

# letra d
ggplot(seguro, aes(x = instrucao, fill = sinistro)) +
  geom_bar(position = "stack")

# letra e
hist(seguro$renda,
      main="Histograma de renda",
      xlab="Renda",
      ylab = "Frequencia",
      col="darkmagenta",
      freq=TRUE)

boxplot(seguro$renda, main="Boxplot renda")

iqr <- IQR(seguro$renda)
q1 <- quantile(seguro$renda, 0.25)
q3 <- quantile(seguro$renda, 0.75)

inf <- q1 - 1.5 * iqr
sup <- q3 + 1.5 * iqr

outliers <- seguro$renda[seguro$renda < inf | seguro$renda > sup]
outliers

install.packages("e1071")
library("e1071")
skewness(seguro$renda)

# letra f
renda <- seguro$renda
mean(renda) # média
median(renda) # mediana
sd(renda) # desvio padrão
coefvari <- (sd(renda)/mean(renda))*100 # coeficiente variacao
quantile(seguro$renda, probs=c(0.25,0.75)) # quartis

# letra g
renda <- seguro$renda
faixas <- cut(
  renda,
  breaks = c(0, 1.9, 3.9, 6.9, Inf),
  labels = c("Até 1.9", "2.0 até 3.9", "4.0 até 6.9", "7.0 ou mais")
)

```

```

tabelafaixas <- table(faixas)
frequencia <- prop.table(tabelafaixas) * 100

tabela <- data.frame(
  faixas = names(tabelafaixas),
  frequencias = as.vector(tabelafaixas),
  frqrelativa = round(as.vector(frequencia), 2)
)
tabela

# letra h
renda <- seguro$renda
ensino <- seguro$instrucao

boxplot(renda ~ ensino,
        main = "Boxplot da Renda por Ensino",
        xlab = "Ensino",
        ylab = "Renda",
        las = 2)

# letra i
renda <- seguro$renda
instrucao <- seguro$instrucao

ensino <- c("fundamental", "médio", "superior")

for (i in ensino) {
  print(i)

  dados <- renda[instrucao == i]

  media <- mean(dados, na.rm = TRUE) #media
  mediana <- median(dados, na.rm = TRUE) #mediana
  q1 <- quantile(dados, 0.25, na.rm = TRUE)#q1
  q3 <- quantile(dados, 0.75, na.rm = TRUE)#q3
  desvio <- sd(dados, na.rm = TRUE)#desvio
  cv <- (desvio / media) * 100#cv

  print(paste("media:", media))
  print(paste("mediana:", mediana))
  print(paste("q1:", q1))
  print(paste("q3:", q3))
  print(paste("desvio padrao:", desvio))
  print(paste("coeficiente de variacao (%):", cv))

  print("-----")
}

# questão 2

```

questão 2

questão 2

Considerando N = 110

N = 110

z = 1,96; d = 6 e o = 14

z = 1.96

d = 6

o = 14

$n0 = ((z^2)*(o^2))/(d^2)$

$n = n0/(1 + n0/N)$

n

questão 3

questão 3

questão 3

media amostral

idades = seguro\$idade

idadessorteadas <- sample(idades, size=18, replace=FALSE)

idadessorteadas

erro <- 6

media <- mean(idadessorteadas)

inf <- media - erro

sup <- media + erro

print(paste("media:", media))

print(paste("erro inferior:", inf))

print(paste("erro superior:", sup))

#media geral

idades = seguro\$idade

erro <- 6

media <- mean(idades)

inf <- media - erro

sup <- media + erro

print(paste("media:", media))

print(paste("erro inferior:", inf))

print(paste("erro superior:", sup))