

Curso Livre I

Aluno: Manuel Ferreira Junior

Matricula: 20180008601

Modulo II

Tarefa 01

Pacotes para pré-instalação e carregamento.

```
install.packages('descr')
install.packages('fdth')
library(descr)
library(fdth)
```

Carregamento dos dados e tratamento

```
url = '/home/manuel/CursoLivre/Curso01/Modulo II/data/trabalhadores.txt'
tbr = read.delim(url)
db = data.frame(tbr)

empregado = db$empregado
salario = db$salario
idade = db$idade
situacao = db$situacao
```

Função que sera usado para calcular a moda nos itens 11 e 12

```
getmode <- function(x){
  z <- table(as.vector(x)); names(z)[z == max(z)]
}
```

1) Transforme a variável situacao em um fator indicando os rótulos de suas classes.

```
> factors <- c('ativo','demitido')
> situacao <- factor(db$situacao,
+                   levels=min(db$situacao):max(db$situacao),
+                   labels=factors)
> situacao
[1] demitido ativo   ativo   ativo   demitido ativo
demitido
[9] ativo   ativo   demitido demitido ativo   ativo   demitido
```

```
demitido
[17] ativo    ativo    ativo    ativo    ativo    ativo    demitido
demitido
[25] demitido ativo    demitido ativo    demitido ativo
Levels: ativo demitido
```

2) Construa uma tabela completa de frequências para a variável situacao. Tabela de frequencia completa

Código :

```
> tf.situacao = freq(situacao)
> tf.situacao
situacao
      Frequency Percent
ativo           18      60
demitido         12      40
Total            30     100
```

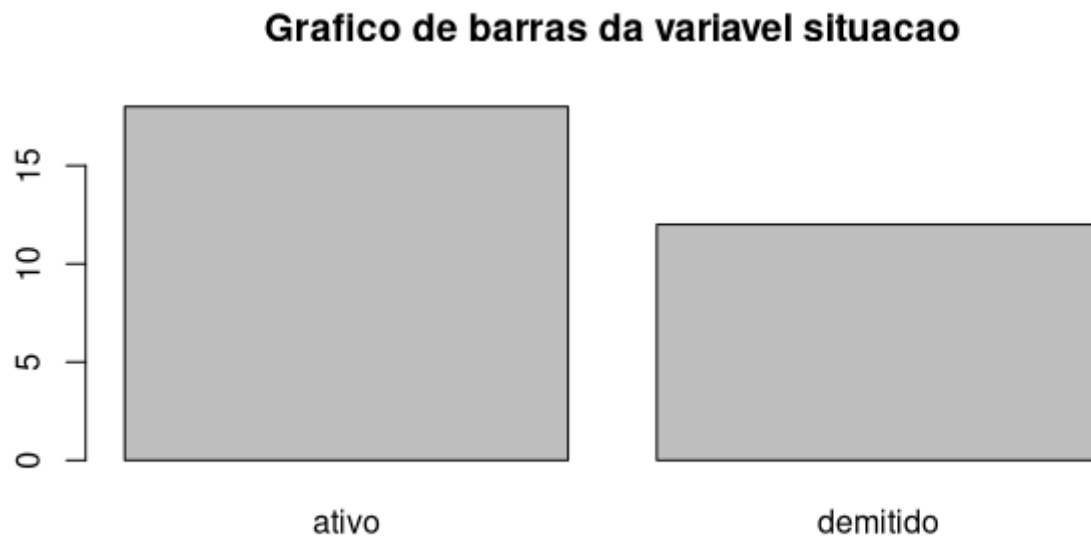
Análise :

Note que 60 dos trabalhadores estão ativos ainda na empresa, e 40\% dos trabalhadores não estão mais ativos.

3) Construa um gráfico de barras ou colunas para a variável situacao. (O grafico ja foi gerado com a tabela da alternativa anterior):

Código :

```
> freq(situacao)
situacao
      Frequency Percent
ativo           18      60
demitido         12      40
Total            30     100
> title('Grafico de barras da variavel situacao')
> # Será retornado logo apos rodar as linhas acima, o grafico.
```

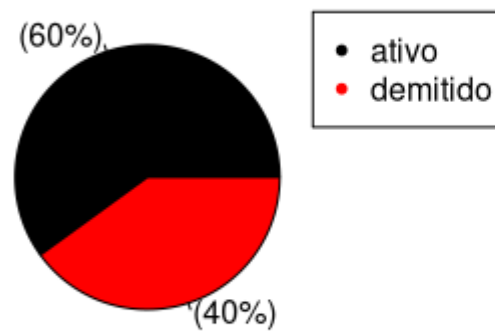
**Analise :**

Perceba que, o número de ativos é superior ao número de demitido dentro da linha de montagem.

4) Construa um gráfico de setores para a variável situacao.**Código :**

```
> porc = round(table(situacao)*100/sum(table(situacao)),2)
> rotulos = paste("(",porc,"%)",sep="")
> pie(table(situacao),labels = rotulos, col=palette())
> legend(1,1,levels(situacao),col=palette(),pch = rep(20,2))
> title('Grafico de setores da variavel situacao')
```

Grafico de setores da variavel situacao



Analise :

Note que o gráfico de setores traz as mesmas informações das alternativas anteriores (2 e 3), ou seja, a informação de que a maioria de trabalhadores em uma determinada linha de montagem possui a situação ativa.

5) Construa uma tabela de frequências por classes com a variável salario.

Código :

```
> tbsal.classes = fdt(salario)
> tbsal.classes
      Class limits f   rf rf(%) cf  cf(%)
[19240.65,23604.667) 6 0.20 20.00  6  20.00
[23604.667,27968.683) 4 0.13 13.33 10  33.33
[27968.683,32332.7) 6 0.20 20.00 16  53.33
[32332.7,36696.717) 1 0.03  3.33 17  56.67
[36696.717,41060.733) 5 0.17 16.67 22  73.33
[41060.733,45424.75) 8 0.27 26.67 30 100.00
```

Analise :

Percebe-se que a maioria dos trabalhadores que exercem função naquela linha de montagem recebem um salario de 41060.733 até 45424.75 dolares anualmente e , apenas um trabalhador daquela linha recebe entre 32332.7 e 36696.717 dólares anualmente.

6) Construa uma tabela de frequência por classes com a variável idade.

Código :

```
> tbid.classes = fdt(idade)
> tbid.classes
  Class limits f   rf rf(%) cf   cf(%)
[18.81,24.933) 4 0.13 13.33  4  13.33
[24.933,31.057) 6 0.20 20.00 10  33.33
[31.057,37.18)  2 0.07  6.67 12  40.00
[37.18,43.303) 8 0.27 26.67 20  66.67
[43.303,49.427) 8 0.27 26.67 28  93.33
[49.427,55.55) 2 0.07  6.67 30 100.00
```

Análise :

Note que a maioria dos trabalhos daqueles trabalhadores são de idades maiores, entre 37.18 anos e 43.303 anos, e também entre 43.303 e 49.427 anos.

7) Construa um boxplot para a variável salario por situação atual de emprego.

Código :

```
> boxplot(salario ~ situacao)
> title('Boxplot da situacao atual do empregado por salario')
> grid()
```



Analise :

Perceba que a média para o salário dos empregados que possuem situação ativa esta em torno de 25000 e 27000 dólares anualmente, enquanto para os que possuem o estado "demitido", a mediana salarial anual é em torno de 40000 e 45000 dólares anual, valor bem superior aos que possuem situação ativa. Note também que o boxplot mostra um outlier salarial para a situação demitido naquela linha de montagem, sendo de algo proximo de 30000 dólares anual.

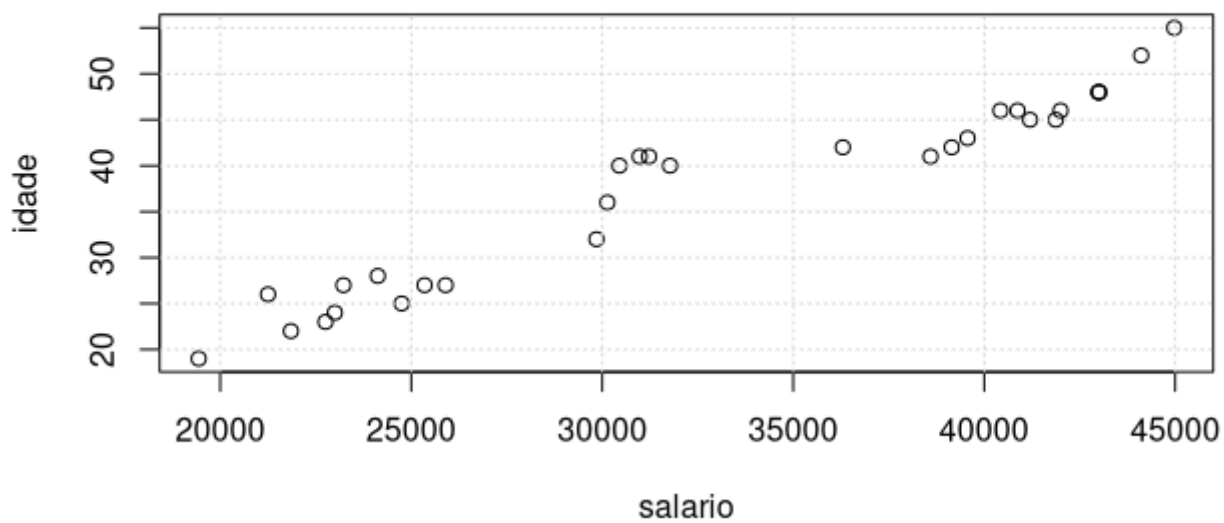
8) Construa um gráfico de dispersão entre as variáveis salario e idade.**Código :**

```
> plot(salario,idade)
> title('Grafico de dispersao entre as variaveis salario e idade')
> grid()
> cor(salario,idade)
[1] 0.9629456
```

Analise :

Note que a relação entre as variáveis salário e idade aparentam ser linearmente dependentes, ao calcular-se uma correlação entre as variáveis obtendo um valor aproximado de 0.9629, aparentemente alta. Perceba que este fato se faz real, ao analisarmos o gráfico, podemos notar que ao aumentar a idade dos empregados dessa linha de montagem, também aumenta o valor do salarial recebido anualmente, da mesma forma, a medida que um salário aumenta, a idade também será maior.

Grafico de dispersao entre as variaveis salario e idade



9) Construa um histograma para a variável salario e insira a curva gaussiana.

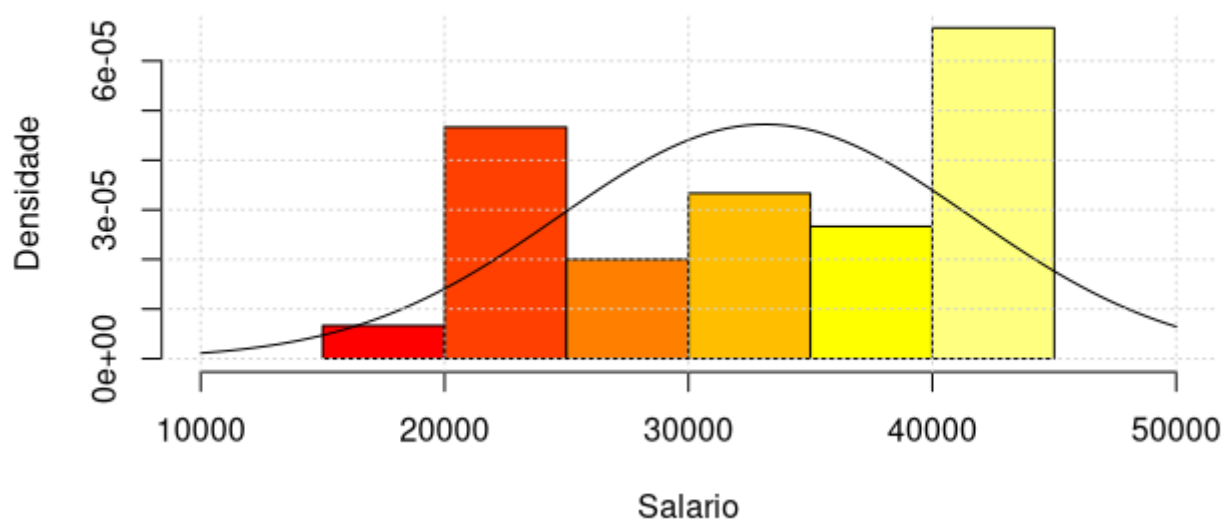
Código :

```
> hist(salario,xlab = "Salario",ylab = "Densidade",  
+      col = heat.colors(6),freq = F,  
+      main = "Histograma da variavel salario",  
+      xlim=c(10000,50000))  
> curve(dnorm(x,mean = mean(salario),  
+            sd = sd(salario)),  
+      add = T)  
> grid()
```

Analise :

Note que a curva gaussiana não se ajusta muito bem para os dados. Percebe que a maioria dos salarios são entr 40000 45000 dólares anuais, e entre 20000 e 25000 dólares anuais, como ja afirmado em alternativas anteriores.

Histograma da variavel salario

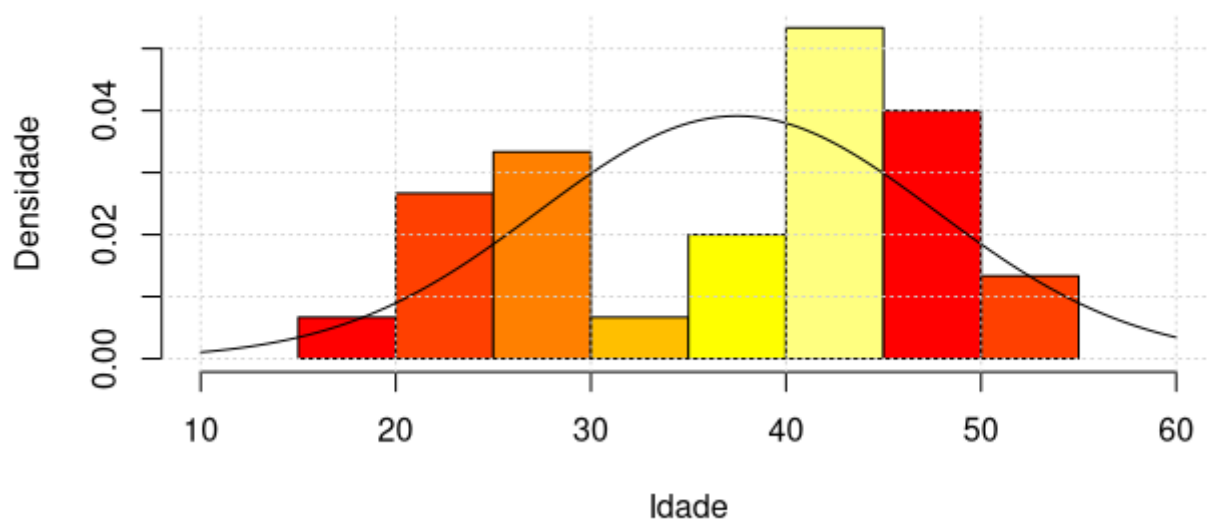


10) Construa um histograma para a variável idade e insira a curva gaussiana.

Código :

```
> hist(idade,xlab = "Idade",ylab = "Densidade",
+       col = heat.colors(6),
+       freq = F,main = "Histograma da variavel idade",
+       xlim=c(10,60))
> curve(dnorm(x,mean = mean(idade),sd = sd(idade)),add = T)
> grid()
```

Histograma da variavel idade



Análise :

Note que a maioria dos trabalhadores que exercem função naquela linha de montagem, ou possuem idade inferior a 30 anos ou superior a 40 anos, sendo sua maioria concentrada na faixa de 30 a 40 anos de idade.

11) Calcule as medidas de tendência central e de dispersão da variável salário.

Código :

```
> c(summary(salario))
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
19435.00 24900.00 31500.00 33137.67 41118.75 44975.00
> c('Mod.' = as.numeric(getmode(salario)))
Mod.
43005
```

Analise :

Note que a maioria dos trabalhadores da linha de montagem recebem 43005 dólares por ano, sendo o menor e maior salário anual de 19435 e 44975, anualmente. Perceba que 75% dos empregados da linha de montagem recebem, anualmente, um salário inferior a 41118.75 dólares, também 50% dos trabalhadores recebem menos de 31500 dólares anualmente, e 50% recebem mais que este valor. Note que a média salarial dos trabalhadores é de 33137.67 dólares anual.

12) Calcule as medidas de tendência central e de dispersão da variável idade.

Código :

```
> c(summary(idade))
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 19.00  27.00  41.00  37.50  45.75  55.00
> c('Mod.' = as.numeric(getmode(idade)))
Mod.1 Mod.2 Mod.3 Mod.4
  27   41   46   48
```

Analise :

Note que a média de idade dos trabalhadores nesta linha de montagem é de 37.5 anos, sendo que 75% dos empregados da mesma, possuem idade superior a 27 anos, com uma idade máxima de 55 anos. Perceba também que 50% dos trabalhadores possuem idade inferior a 41 anos, e 50% superior a 41 anos.

13) Calcule média, mediana e o desvio-padrão da variável idade por situação de emprego.

Código :

```
> mean.idsit = tapply(idade,situacao,mean)
> median.idsit = tapply(idade,situacao,median)
> sd.idsit = tapply(idade,situacao,sd)
> sit.id = data.frame(row.names = c('Média',
+                                'Mediana',
+                                'Desvio-Padrao'),
+                    'ativo' = c(mean.idsit[1],
+                                median.idsit[1],
+                                sd.idsit[1]),
+                    'demitido' = c(mean.idsit[2],
+                                median.idsit[2],
+                                sd.idsit[2]))
> sit.id
```

	ativo	demitido
Média	31.555556	46.416667
Mediana	27.500000	46.000000
Desvio-Padrao	8.325879	4.718596

Análise :

Note que a média da idade dos empregados que possuem situação ativada na linha de montagem é de 31.56 anos aproximadamente, e a média de idade para a situação demitido, é de 46.42 anos aproximadamente. Perceba que a variabilidade das idades é menor para os que possuem situação demitido. Note também que 50% dos ativos, possuem idade inferior a 27.5 anos e 50% superior a 27.5 anos, da mesma forma, notamos que 50% dos demitidos possuem idade superior a 46 anos e 50% superior a esta idade.

14) Construa uma tabela entre as variáveis salario e situacao, utilizando o comando compmeans.

Código :

```
> tb.comp = compmeans(salario,situacao)
> tb.comp
Mean value of "salario" according to "situacao"
      Mean  N Std. Dev.
ativo   27798.89 18  6069.148
demitido 41145.83 12  3824.089
Total   33137.67 30  8447.938
> title('Boxplot do salario por situacao do individuo')
> grid()
```

Analise :

Perceba que a média salarial para quem possui situação demitido é quase o dobro da média salarial dos que possuem situação de ativo, sendo grupos de tamanho 12 e 18 trabalhadores, respectivamente. Perceba também que a variabilidade continua sendo menor para o grupo que possui situação de demitido.