

# Análise de Dados - UFPE/2019 - Lista 5

Antonio Fernandes

14 de maio de 2019

header-includes:

## Conteúdo

Apresentação . . . . .	1
Questão 5 - Lista I . . . . .	1
Questão 6 - Lista I . . . . .	3
Questão 7 - Lista I . . . . .	4
Questão 8 - Lista I . . . . .	6
Questão 9 - Lista I . . . . .	6

## Apresentação

Este documento apresenta as respostas dos exercícios realizados à mão presentes nas listas 1 e 2. Os scripts estão no repositório do GitHub.

## Questão 5 - Lista I

Primeiramente foi criado o Dataframe com as informações necessárias:

```
Emp <- data.frame("EMP" = 1:15, "MES" = c(8,9,4,5,3,6,8,6,6,8,5,5,6,4,4),  
                  "SET" = c('C','C','I','I','I','C','C','I','I','C','C','I','C','I','I'),  
                  "TAM" = c('G','M','G','M','M','P','G','M','P','M','P','P','M','M','G'))
```

Após isso, o banco foi dividido entre Comércio e Indústria:

```
C <- Emp[ which(Emp$SET == "C"),]  
I <- Emp[ which(Emp$SET == "I"),]
```

Em seguida, foram calculadas a média, moda e mediana de cada setor:

```
#Media e mediana de cada grupo
```

```
mean(C$MES)
```

```
## [1] 7.142857
```

```
mean(I$MES)
```

```
## [1] 4.625
```

```
median(C$MES)
```

```
## [1] 8
```

```
median(I$MES)
```

```
## [1] 4.5
```

```
#Desvio Padrão de cada grupo
```

```
sd(C$MES)
```

```
## [1] 1.46385
```

```
sd(I$MES)
```

```
## [1] 1.06066
```

Onde percebe-se que a média, mediana e desvio padrão dos meses com crescimento é maior no comércio.

O próximo passo é indentificar o número máximo de meses com crescimento para a empresa receber um incentivo fiscal. Nesse caso, serão as empresas com meses menores que o 25 decil:

```
fivenum(Emp$MES)
```

```
## [1] 3.0 4.5 6.0 7.0 9.0
```

Nesse caso, o máximo é 4.5 meses. Por fim, é necessário verificar as estatísticas descritivas de acordo com o porte da empresa:

```
G <- Emp[ which(Emp$TAM == "G"),] ##Tamanho grande  
M <- Emp[ which(Emp$TAM == "M"),] ##Tamanho médio  
P <- Emp[ which(Emp$TAM == "P"),] ##Tamanho pequeno
```

```
median(G$MES)
```

```
## [1] 6
```

```
median(M$MES)
```

```
## [1] 6
```

```
median(P$MES)
```

```
## [1] 5.5
```

```
mean(G$MES)
```

```
## [1] 6
```

```
mean(M$MES)
```

```
## [1] 5.857143
```

```
mean(P$MES)
```

```
## [1] 5.5
```

```
sd(G$MES)
```

```
## [1] 2.309401
```

```
sd(M$MES)
```

```
## [1] 2.115701
```

```
sd(P$MES)
```

```
## [1] 0.5773503
```

A mediana de meses com crescimento é a mesma para empresas de grande e médio porte (6) e um pouco menor para empresas de pequeno porte (5.5). Já em relação a média, as empresas de grande porte apresentam uma quantidade de meses com crescimento um pouco maior que as empresas de médio e pequeno porte. Do mesmo modo, no tocante ao desvio padrão, as empresas grandes apresentam uma maior variação nos meses com crescimento do que as empresas de médio e pequeno porte.

## Questão 6 - Lista I

Primeiramente é utilizado o comando `data.frame` para criar um data frame contendo as informações necessárias:

```
Inv <- data.frame("CID" = c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J"),  
                  "INV" = c(26,16, 14,10, 19, 15, 19, 16, 19, 18))
```

O cálculo do investimento médio ocorre da seguinte maneira:

```
mean(Inv$INV)
```

```
## [1] 17.2
```

Ou seja, o investimento médio é de 17.2

Agora será calculado quais cidades receberão o programa especial:

```
s <- sd(Inv$INV)

mean(Inv$INV) - (2*s) ##Identificando valor para programa especial
```

```
## [1] 8.830744
```

Cidades com menos de 8.83 em investimentos receberão o programa especial. Agora, será verificado qual o valor mínimo e máximo de investimento básico e quais cidades estão neste critério.

```
mean(Inv$INV) - (2*s) #Valor minimo investimento básico
```

```
## [1] 8.830744
```

```
mean(Inv$INV) + (2*s) #Valor maximo investimento basico
```

```
## [1] 25.56926
```

```
invba <- Inv[ which(Inv$INV < 25.56
                    & Inv$INV > 8.83), ] #Selecionando casos com investimento básico

mean(invba$INV) #media investimento basico
```

```
## [1] 16.22222
```

Podemos perceber que a média de investimento básico acaba sendo menor do que a média de investimento total ( $16.22 < 17.2$ ).

## Questão 7 - Lista I

Primeiro, vamos criar o banco de dados contendo as informações dos estímulos visuais:

```
Est <- data.frame("IND" = 1:20, "A" = c(55,2,13,11,23,2,15,12,14,28,12,45,19,30,16,12,7,13,1,7),
                  "B" = c(20,7,6,5,3,25,5,3,3,10,8,5,1,35,9,8,12,2,26,NA))
```

O próximo passo é obter as estatísticas descritivas do banco:

```
mean(Est$A) ##Media estimulo A = 16.85
```

```
## [1] 16.85
```

```
mean(Est$B, na.rm = TRUE) ##Media estimulo B = 10.16
```

```
## [1] 10.15789
```

```
median(Est$A) ##Mediana estimulo A = 13
```

```
## [1] 13
```

```
median(Est$B, na.rm = TRUE) ##Mediana estimulo B = 7
```

```
## [1] 7
```

```
sd(Est$A) ##Desvio padrao estimulo A = 13.80
```

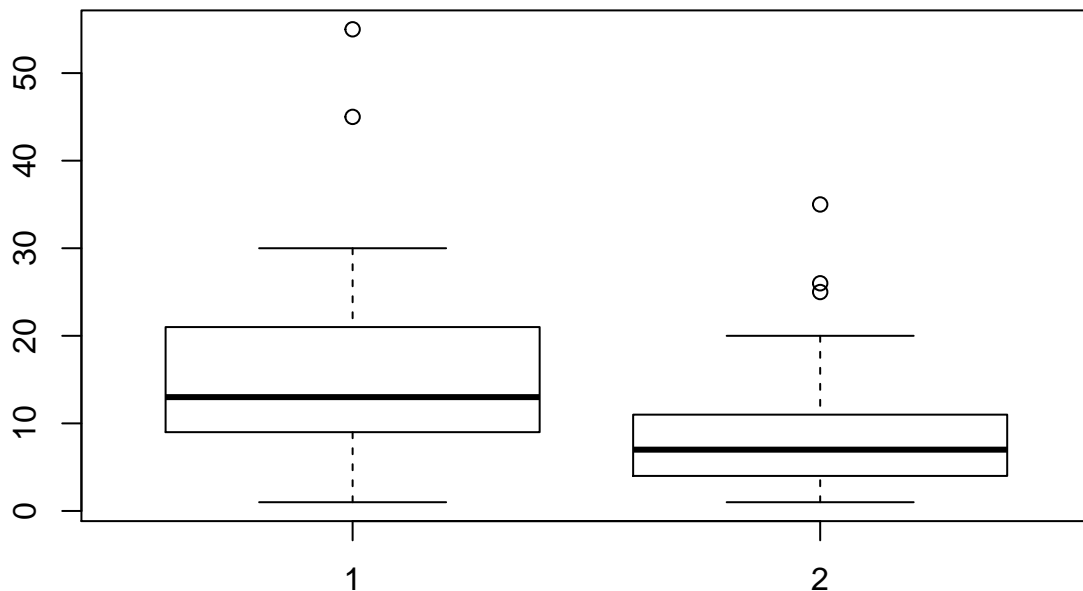
```
## [1] 13.80418
```

```
sd(Est$B, na.rm = TRUE) ##Desvio padrao estimulo B = 9.45
```

```
## [1] 9.459053
```

Os resultados mostram que a média do tempo de reação do estímulo A é maior do que a média de reação do estímulo B. Do mesmo modo, a média do estímulo A é 13 enquanto que em relação ao estímulo B esse valor é de 7. Ao analisar o desvio padrão dos dois estímulos, o desvio padrão do estímulo A é de 13.80 enquanto que o do estímulo B é de 9.45, sendo possível inferir que o estímulo B possui uma menor variação no tempo de resposta.

Por fim, é possível observarmos a distribuição das variáveis por meio de um boxplot:



Podemos visualizar que ambas as distribuições apresentam alguns valores destoantes, que no estímulo A estão acima de 40 (55 e 45) e no estímulo B estão acima de 20 (35, 26 e 25).

## Questão 8 - Lista I

Na questão 8, também precisamos iniciar criando o banco de dados para realizar as análises:

```
fam <- data.frame("FAM" = c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J"), "REN" =  
                  c(12,16,18,20,28,30,40,48,50,54), "SAU" = c(7.2,7.4,7,6.5,6.6,6.7,6,5.6,6,5.5))
```

Após isso, podemos obter as informações solicitadas:

```
mean(fam$REN) #Media renda familia
```

```
## [1] 31.6
```

```
mean(fam$SAU) #Media percentual gasto com saude
```

```
## [1] 6.45
```

```
sd(fam$REN) #sd renda familia
```

```
## [1] 15.42869
```

```
sd(fam$SAU) #sd percentual gasto com saude
```

```
## [1] 0.6570134
```

```
cov(fam$REN, fam$SAU) #cov entre as duas variaveis
```

```
## [1] -9.533333
```

```
cor(fam$REN, fam$SAU) #correlacao entre as duas variaveis
```

```
## [1] -0.9404625
```

Os resultados mostram que a média de renda da família é de 31.6 e a média do gasto percentual com saúde é de 6.45. Já em relação ao desvio padrão, o valor obtido em relação a renda da família foi de 15.42 e o gasto percentual com saúde foi de 0.66. A covariância entre as duas variáveis foi -9.53 e a correlação foi de -0.94.

Ou seja, a correlação mostrou uma relação inversamente proporcional entre as duas variáveis, indicando que o aumento de uma leva a queda da outra. Além disso, o valor obtido muito próximo de 1 mostra que o grau de associação entre as variáveis é forte.

## Questão 9 - Lista I

Na questão 9, primeiro será criado um banco contendo as informações dos alunos:

```
not <- data.frame("ALU" = c("A", "B", "C", "D", "E", "F", "G", "H", "I"), "P1" =  
                  c(7.5,8.2,8.5,8.7,8.8,9.1,9.2,9.3,10), "P2" = c(8.2, 8,8.3,  
                                                                8.5,9.4,9.6,9,9.3,9.7))
```

Com isso, podemos realizar uma análise de correlação entre as notas da prova I e as notas da prova II:

```
cor(not$PI, not$P2)
```

```
## [1] 0.8301592
```

O valor obtido na correlação de 0.83 indica uma associação forte entre as variáveis, mostrnado que estas estão altamente correlacionadas. Ou seja, uma nota alta na prova I indica uma nota alta também na prova II.