

Análise de Dados - UFPE/2019 - Lista 5

Antonio Fernandes

14 de maio de 2019

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:

```
#Média 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
```

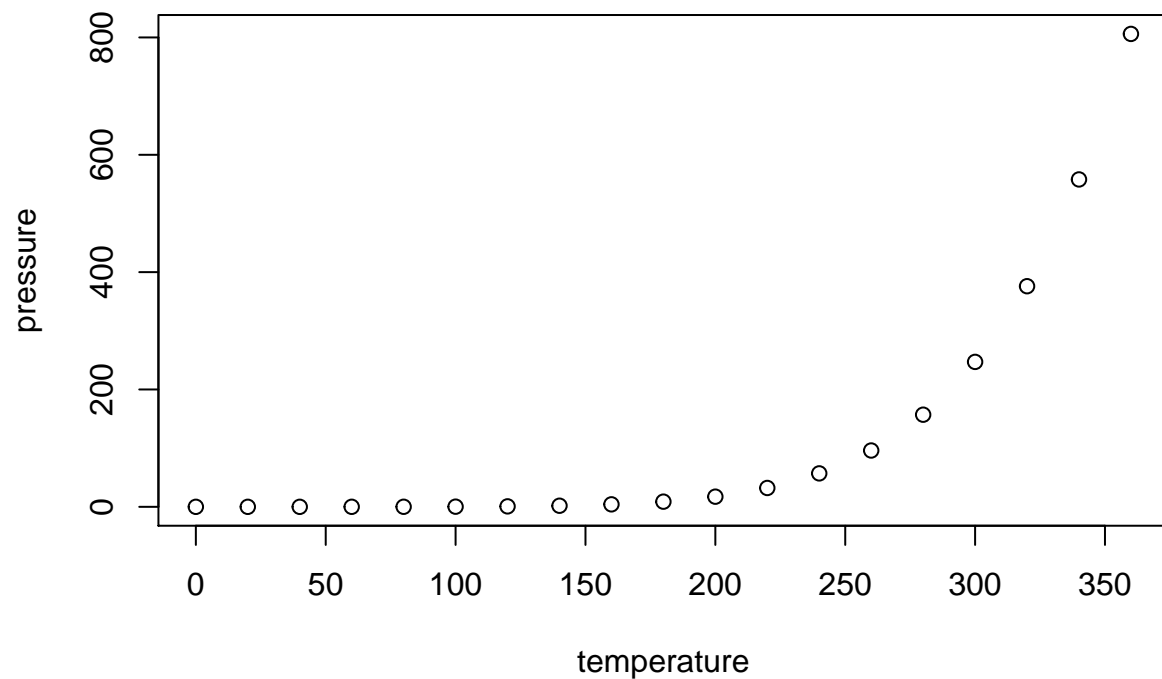
```
invba <- Inv[ which(Inv$INV < 25.56  
                  & Inv$INV > 8.83), ] ##Selecionando casos com investimento básico  
mean(invba$INV) #media investimento basico
```

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

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

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.