



## HOMEWORK I

NOME COMPLETO:

NUMERO DE MATRICULA:

### QUESTÃO 1

As emissões diárias de um gás poluente de uma planta industrial foram registradas 80 vezes, em uma determinada unidade de medida. Os dados obtidos estão apresentados na Tabela 1.

|      |      |      |      |      |      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|------|------|------|------|------|
| 15.8 | 22.7 | 26.8 | 19.1 | 18.5 | 14.4 | 8.3  | 25.9 | 26.4 | 9.8  | 21.9 | 10.5 |
| 17.3 | 6.2  | 18.0 | 22.9 | 24.6 | 19.4 | 12.3 | 15.9 | 20.1 | 17.0 | 22.3 | 27.5 |
| 23.9 | 17.5 | 11.0 | 20.4 | 16.2 | 20.8 | 20.9 | 21.4 | 18.0 | 24.3 | 11.8 | 17.9 |
| 18.7 | 12.8 | 15.5 | 19.2 | 13.9 | 28.6 | 19.4 | 21.6 | 13.5 | 24.6 | 20.0 | 24.1 |
| 9.0  | 17.6 | 25.7 | 20.1 | 13.2 | 23.7 | 10.7 | 19.0 | 14.5 | 18.1 | 31.8 | 28.5 |
| 22.7 | 15.2 | 23.0 | 29.6 | 11.2 | 14.7 | 20.5 | 26.6 | 13.3 | 18.1 | 24.8 | 26.1 |
| 7.7  | 22.5 | 19.3 | 19.4 | 16.7 | 16.9 | 23.5 | 18.4 |      |      |      |      |

Tabela 1: Emissões diárias de gás poluente (questão 1).

1. Calcule as medidas de tendência central (média, mediana e moda) e as medidas de dispersão (amplitude, variância, desvio padrão e coeficiente de variação) para o conjunto de dados da Tabela 1. Interprete os resultados.
2. Construa um histograma e um boxplot para os dados de emissões. Os dados parecem estar simetricamente distribuídos? Existem valores atípicos?
3. Determine os quartis (Q1, Q2, Q3) e o intervalo interquartil (IQR). Utilize esses valores para reforçar sua análise sobre a presença de valores atípicos.
4. Suponha que o limite máximo aceitável diário para as emissões seja de 25 unidades. Qual a proporção de dias em que a planta excedeu esse limite? O comportamento geral das emissões estaria em conformidade com esse padrão regulatório?

## SOLUÇÃO DA QUESTÃO 1

Explique brevemente a base teórica, as principais etapas de sua solução e os resultados. Um exemplo de gráficos é reportado na figura 1.

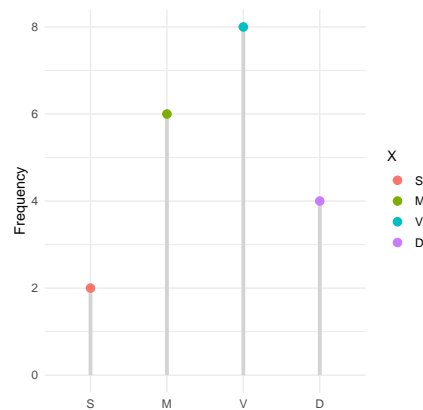


Figura 1: Exemplo de gráfico de linha

O código é reportado, por exemplo, no Listado 1 ou no Listado da apêndix A.

Listado 1: Solution of exercise 1

```
rm(list=ls())           # clean the working space
graphics.off()          # close all the graphic windows
getwd()                 # verify the current working directory
setwd('path/TI0111/my_folder') # set your working directory
```

## QUESTÃO 2

Uma empresa italiana recebeu 20 currículos de cidadãos italianos e estrangeiros na seleção de pessoal qualificado para o cargo de gerente de relações exteriores. A tabela 2 reporta as informações consideradas relevantes na seleção: a idade, a nacionalidade, o nível mínimo de renda desejada (em milhares de euros), os anos de experiência no trabalho.

1. Calcule a média, mediana e desvio padrão para as variáveis idade, renda desejada e anos de experiência. O que você pode inferir a partir desses valores sobre o perfil típico dos candidatos?
2. Agrupe os candidatos por nacionalidade e calcule a renda média desejada e os anos médios de experiência para cada grupo. Qual nacionalidade apresenta a maior renda média desejada? Qual grupo aparenta ser o mais experiente?
3. Existe correlação entre anos de experiência e renda desejada? Utilize ferramentas visuais apropriadas (por exemplo, gráfico de dispersão) e calcule o coeficiente de correlação de Pearson. Interprete o resultado.

|    | Idade | Nacionalidade | Renda | Experiência |
|----|-------|---------------|-------|-------------|
| 1  | 28    | Italiana      | 2.3   | 2           |
| 2  | 34    | Inglesa       | 1.6   | 8           |
| 3  | 46    | Belga         | 1.2   | 21          |
| 4  | 26    | Espanhola     | 0.9   | 1           |
| 5  | 37    | Italiana      | 2.1   | 15          |
| 6  | 29    | Espanhola     | 1.6   | 3           |
| 7  | 51    | Francesa      | 1.8   | 28          |
| 8  | 31    | Belga         | 1.4   | 5           |
| 9  | 39    | Italiana      | 1.2   | 13          |
| 10 | 43    | Italiana      | 2.8   | 20          |
| 11 | 58    | Italiana      | 3.4   | 32          |
| 12 | 44    | Inglesa       | 2.7   | 23          |
| 13 | 25    | Francesa      | 1.6   | 1           |
| 14 | 23    | Espanhola     | 1.2   | 0           |
| 15 | 52    | Italiana      | 1.1   | 29          |
| 16 | 42    | Alemana       | 2.5   | 18          |
| 17 | 48    | Francesa      | 2.0   | 19          |
| 18 | 33    | Italiana      | 1.7   | 7           |
| 19 | 38    | Alemana       | 2.1   | 12          |
| 20 | 46    | Italiana      | 3.2   | 23          |

Tabela 2: Informações na seleção da empresa italiana (questão 2).

- Suponha que a empresa queira priorizar candidatos com pelo menos 10 anos de experiência e renda desejada inferior a 2,0 (mil euros). Quantos candidatos atendem a ambos os critérios? Liste suas nacionalidades e idades.
- Construa gráficos que permitam visualizar a distribuição da idade e da renda desejada, separados por nacionalidade. Utilize histogramas, box-plots ou gráficos de barras, e comente as principais diferenças observadas entre os grupos.

### SOLUÇÃO DA QUESTÃO 2

Explique brevemente a base teórica, as principais etapas de sua solução e os resultados. Reporte o código conforme os exemplos anteriores.

### QUESTÃO 3

O conjunto de dados em anexo, `HW1_bike_sharing.csv`<sup>1</sup>, refere-se ao processo de compartilhamento de bicicletas em uma cidade dos Estados Unidos. O conjunto contém as colunas descritas na Tabela 3. A variável `season` inclui as quatro estações do hemisfério

<sup>1</sup> Os dados estão disponíveis no material do homework.

norte: primavera, verão, outono e inverno. A variável **weathersit** representa quatro condições meteorológicas: ‘Céu limpo’, ‘Nublado’, ‘Chuva fraca’, ‘Chuva forte’. A variável **temp** é a temperatura normalizada em graus Celsius, ou seja, os valores foram divididos por 41 (valor máximo).

| TAG               | DESCRIÇÃO                       |
|-------------------|---------------------------------|
| <b>instant</b>    | Índice de registro              |
| <b>dteday</b>     | Data da observação              |
| <b>season</b>     | Estação do ano                  |
| <b>weathersit</b> | Condições meteorológicas        |
| <b>temp</b>       | Temperatura em °C (normalizada) |
| <b>casual</b>     | Número de usuários casuais      |
| <b>registered</b> | Número de usuários registrados  |

Tabela 3: Variáveis do conjunto **HW1\_bike\_sharing** (questão 3).

1. Carregue o conjunto de dados **HW1\_bike\_sharing.csv** no R. Classifique as variáveis quanto ao tipo (categórica ou numérica), identifique o número total de observações e as datas de início e fim da amostra.
2. Calcule medidas de tendência central (média, mediana) e os quartis para cada característica numérica relevante. Apresente os resultados em uma tabela com título apropriado. Comente os principais pontos.
3. Atribua os níveis correspondentes às variáveis **season** e **weathersit**. Construa gráficos de barras para ambas. Qual estação do ano apresenta maior número de usuários? O uso de bicicletas depende da estação? Qual é a condição climática mais favorável para o uso do sistema?
4. Calcule o número total de usuários por dia, somando **casual** e **registered**. Converta a variável **temp** para temperatura real (multiplicando por 41). Em seguida, construa os gráficos de séries temporais para temperatura e número total de usuários. Essas séries apresentam tendência semelhante?

### SOLUÇÃO DA QUESTÃO 3

Explique brevemente a base teórica, as principais etapas de sua solução e os resultados. Reporte o código conforme os exemplos anteriores.

## CÓDIGOS A

### Listado 2: Solution of exercise [1](#)

```
rm(list=ls())           # clean the working space
graphics.off()          # close all the graphic windows
getwd()                 # verify the current working directory
setwd('path/TI0111/my_folder') # set your working directory
```

## CÓDIGOS B

### Listado 3: Solution of exercise [2](#)

```
rm(list=ls())           # clean the working space
graphics.off()          # close all the graphic windows
getwd()                 # verify the current working directory
setwd('path/TI0111/my_folder') # set your working directory
```