



escola
britânica de
artes criativas
& tecnologia

Módulo | Análise de Dados: Fundamentos de Estatística

Caderno de **Exercícios**

Professor [André Perez](#)

Tópicos

1. Média e Variância;
2. Ordem e Posição;
3. Correlação.

Exercícios

1. Tráfego de São Paulo

Neste exercício, vamos continuar a analisar os dados de mobilidade urbana da cidade de São Paulo. A base de dados está neste [link](#) e é uma cópia do dado original, presente neste [link](#). A base de dados contém a quantidade de acidentes ocorridos na cidade entre 14/12/09 e 18/12/09, das 07:00h às 20:00h, agregados em intervalos de 30 minutos.

```
In [ ]: !wget -q "https://raw.githubusercontent.com/andre-marcos-perez/ebac-course-
```

```
In [ ]: import pandas as pd
```

```
In [ ]: df = pd.read_csv('traffic.csv', sep=';')
```

```
In [ ]: df.head()
```

- **Exemplo:** (para lembrar)

O código abaixo extrai a 13ª linha do arquivo que representa a 14ª meia hora do dia 14/12/09, contadas a partir das 07:00h, ou seja, todos os incidentes ocorridos na cidade no dia em questão entre as 13:30h e 14:00h. Vemos, por exemplo, que dois ônibus e dois caminhões ficaram parados nas ruas da cidade, atrapalhando o tráfego.

```
In [ ]: df.iloc[[13]]
```

1.1. Agregação

Nesta primeira atividade, você deve gerar um array NumPy por dia. Para cada array você deve somar todos os incidentes que aconteceram naquela meia hora. Sendo assim, cada array deve ter 27 posições, cada qual com a soma dos incidentes daquela meia hora.

Dica: Você deve remover a primeira e a última coluna.

Dica: Os arrays devem ter os seguintes valores:

```
14: [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 5, 2, 0, 0, 0, 0, 0, 1,
0, 0, 0, 2, 2, 3]
15: [3, 0, 0, 0, 2, 4, 2, 1, 3, 0, 2, 0, 2, 3, 2, 2, 4, 7, 6, 3, 0,
1, 1, 5, 4, 6, 0]
16: [1, 1, 2, 4, 4, 5, 3, 2, 4, 2, 1, 1, 3, 6, 2, 0, 2, 8, 4, 6, 4,
1, 2, 7, 15, 4, 0]
17: [4, 1, 1, 0, 0, 0, 2, 3, 2, 7, 2, 1, 7, 3, 6, 4, 3, 2, 2, 3, 2,
5, 2, 0, 1, 4, 15]
18: [10, 1, 1, 5, 5, 7, 3, 4, 3, 1, 2, 3, 2, 2, 6, 2, 2, 4, 2, 3,
0, 3, 3, 2, 3, 6, 0]
```

```
In [ ]: # resposta da questão 1.1
```

1.2. Métricas

Para cada array você deve calcular as seguintes métricas:

- **média;**
- **desvio padrão.**

```
In [ ]: # resposta da questão 1.2
```

1.3. Interpretação

Baseado nos resultados da questão 1.2, responda:

Qual dia apresenta a maior média de acidentes por meia hora?

Resposta:

Qual dia apresenta a menor variação de acidentes por meia hora?

Resposta:
