

# Basic Python - Basic Data Analysis

*Hoàng-Nguyên Vũ*

## 1 Correlation (Tương quan)

Correlation là một phép đo thống kê cho biết mức độ và hướng của mối quan hệ giữa hai biến số. Khi hai biến số có mối quan hệ tương quan, giá trị của một biến có thể ảnh hưởng đến giá trị của biến còn lại.

- **Positive Correlation (Tương quan dương):** Khi giá trị của một biến tăng thì giá trị của biến kia cũng tăng.
- **Negative Correlation (Tương quan âm):** Khi giá trị của một biến tăng thì giá trị của biến kia giảm.
- **No Correlation (Không tương quan):** Không có mối quan hệ rõ ràng giữa hai biến.

## 2 Correlation Coefficient (Hệ số tương quan)

Hệ số tương quan là một giá trị số từ -1 đến 1 cho biết mức độ của mối quan hệ tương quan:

- **1:** Mối quan hệ tương quan hoàn hảo dương.
- **0:** Không có mối quan hệ tương quan.
- **-1:** Mối quan hệ tương quan hoàn hảo âm.

**Công thức tính hệ số tương quan Pearson (Pearson Correlation Coefficient)**

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$$

Trong đó:

- $X$  và  $Y$  là hai biến số.
- $\bar{X}$  và  $\bar{Y}$  là giá trị trung bình của  $X$  và  $Y$ .

## 3 Bài tập: Tính hệ số tương quan giữa các feature trong tập dữ liệu advertise

1. Đọc file dữ liệu advertising, tại đây.
2. Hãy tính hệ số tương quan giữa các features có trong tập dataset này.

## Hướng dẫn:

```

1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5
6 # Đọc dataset
7 data = pd.read_csv("advertising.csv")
8
9 def correlation(x, y):
10     ## Your code here ##
11
12 # Example usage:
13 x = data['TV']
14 y = data['Radio']
15 corr_xy = correlation(x, y)
16 print(f"Correlation between TV and Sales: {corr_xy}") ## Output: 0.0548

```

### 3.1 Bài tập mở rộng

Từ giá trị correlation coefficient ở trên, hãy vẽ biểu đồ heatmap để biểu hiện sự tương quan.

## Hướng dẫn:

```

1 ## Calculate coereleation here ##
2 plt.figure(figsize=(10,8))
3 sns.heatmap(#Your result coereleation here#, annot=True, fmt=".2f",
4             linewidth=.5)
5 plt.show()

```

