

# 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  $Y$ .

## Bài Tập 1: Mối Quan Hệ Giữa Quảng Cáo và Doanh Thu

Một công ty muốn phân tích xem có mối quan hệ giữa chi tiêu quảng cáo và doanh thu hay không. Bạn có dữ liệu như sau:

- $X = [10, 20, 30, 40, 50, 60, 70, 80, 90, 100]$  (triệu VND, chi tiêu quảng cáo).
- $Y = [12, 25, 28, 35, 48, 55, 70, 78, 90, 102]$  (triệu VND, doanh thu).

### Yêu cầu:

1. Tính hệ số tương quan Pearson giữa chi tiêu quảng cáo và doanh thu.
2. Vẽ biểu đồ scatter plot.
3. Đưa ra nhận xét về mối quan hệ giữa hai biến.

## Bài Tập 2: Ảnh Hưởng Của Thời Tiết Đến Doanh Số Cà Phê

Một quán cà phê muốn kiểm tra xem nhiệt độ ngoài trời có ảnh hưởng đến doanh số bán cà phê hay không. Dữ liệu:

- $X = [30, 28, 25, 30, 29, 27, 26, 24, 23, 22]$  (nhiệt độ, độ C).
- $Y = [50, 55, 60, 65, 70, 75, 80, 85, 90, 95]$  (số ly cà phê bán được).

### Yêu cầu:

1. Tính hệ số tương quan Pearson giữa nhiệt độ và số ly cà phê bán được.
2. Vẽ scatter plot để phân tích xu hướng.
3. Đưa ra nhận xét.