Phân tích đơn biến và hai biến trong khám phá dữ liệu

# 1. Phân tích đơn biến (Univariate Analysis)

Phân tích đơn biến tập trung phân tích một biến duy nhất trong tập dữ liệu, để hiểu đặc điểm phân phối, xu hướng trung tâm và mức độ phân tán.  
Ví dụ: kiểm tra phân phối chiều cao của sinh viên trong một lớp.

# 2. Phân tích hai biến (Bivariate Analysis)

Phân tích hai biến xem xét mối quan hệ giữa hai biến (có thể là định lượng - định lượng, định lượng - định tính, hoặc định tính - định tính).  
Ví dụ: phân tích mối quan hệ giữa chiều cao và cân nặng.

# 3. Các thước đo thống kê trong phân tích đơn biến

- Trung bình (Mean)  
- Trung vị (Median)  
- Mode (giá trị xuất hiện nhiều nhất)  
- Phương sai (Variance), Độ lệch chuẩn (Standard Deviation)  
- Min, Max, Range, Quartile, IQR  
- Skewness (độ lệch), Kurtosis (độ nhọn)

# 4. Xác định mối quan hệ giữa hai biến

- Hệ số tương quan (Pearson/Spearman/Kendall)  
- Phân tích hồi quy (Regression)  
- Bảng chéo (Cross-tabulation, Chi-square test)  
- Biểu đồ trực quan hóa (Scatter plot, Boxplot, Violin plot, Heatmap)

# 5. Sự khác biệt giữa Tương quan và Hiệp biến

- Covariance (hiệp biến): đo mức độ hai biến thay đổi cùng nhau, phụ thuộc vào đơn vị đo.  
- Correlation (tương quan): chuẩn hóa từ hiệp biến, giá trị [-1, 1], dễ so sánh, không phụ thuộc đơn vị đo.

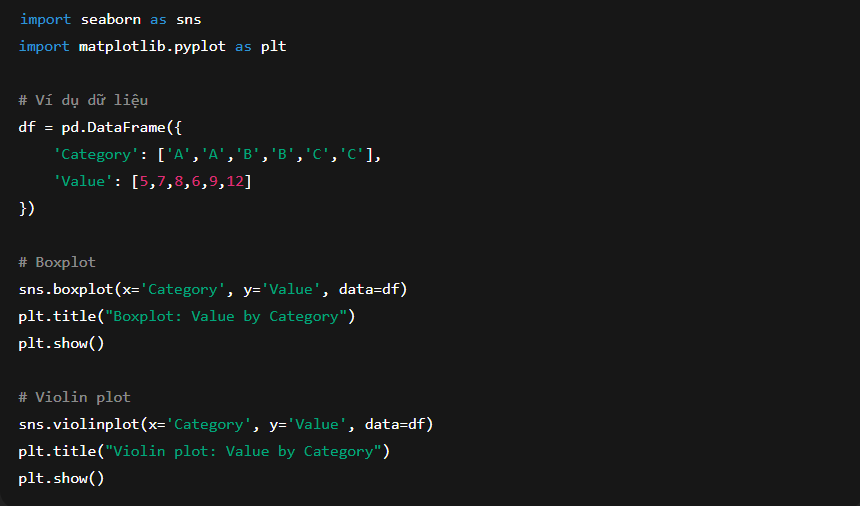
# 6. Khi nào dùng biểu đồ trực quan hóa

- Đơn biến: Histogram, Boxplot, Density plot  
- Hai biến: Scatter plot, Heatmap, Boxplot/Violin plot, Bar chart

# 7. Code mẫu trực quan hóa

# 

1. **Trực quan hóa mối quan hệ biến số & biến phân loại (Boxplot/Violin Plot)**



Bảng so sánh nhanh giữa **Phân tích đơn biến** và **Phân tích hai biến:**

|  |  |  |
| --- | --- | --- |
| Đặc điểm | Univariate Analysis | Bivariate Analysis |
| Số biến | 1 biến duy nhất | 2 biến cùng lúc |
| Mục đích | Mô tả đặc điểm phân phối | Khám phá mối quan hệ |
| Thước đo | Mean, Median, Mode, Std, Skewness, Kurtosis | Correlation, Covariance, Regression, Chi-square |
| Biểu đồ | Histogram, Boxplot, Density plot | Scatter plot, Heatmap, Boxplot, Violin plot |
| Ứng dụng | Phát hiện ngoại lai, kiểm tra phân phối | Tìm mối liên hệ, xu hướng |
| Ví dụ | Phân tích chiều cao của sinh viên | Phân tích mối quan hệ giữa chiều cao và cân nặng |