

Tuần 6: Biểu đồ nâng cao & Thao tác dữ liệu

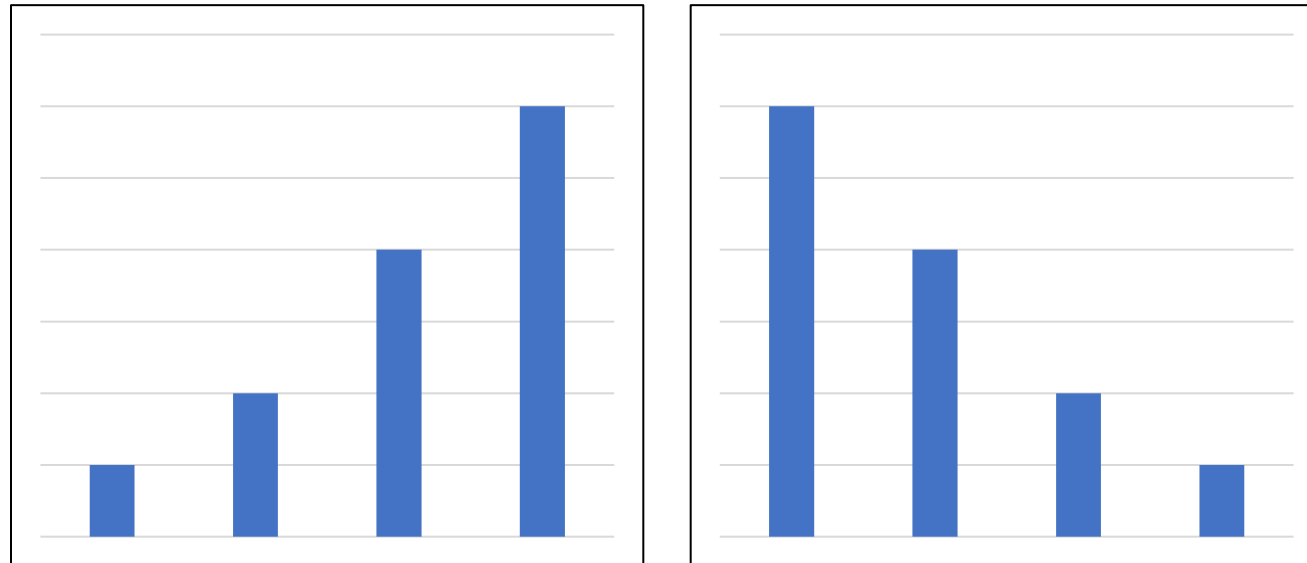
Giảng viên: ThS. Lê Minh Tân

Nội dung

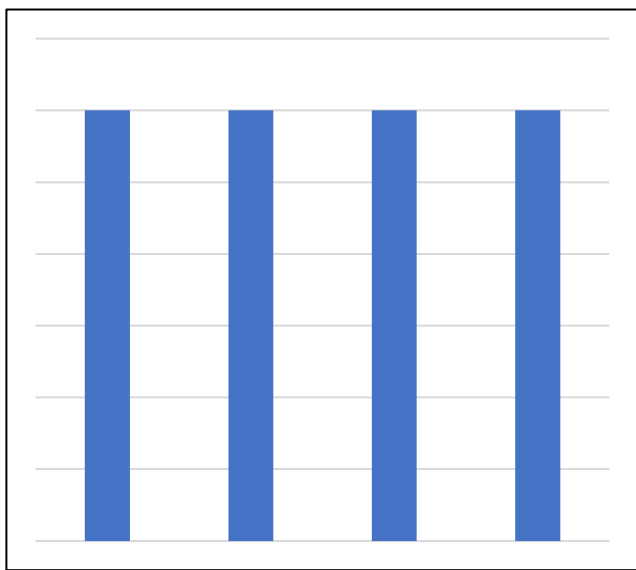
- I. Đường xu hướng, đường mục tiêu
- II. Thông tin dữ liệu và thao tác dữ liệu
- III. Trực quan hóa trên Zeppelin

I. Đường xu hướng, đường mục tiêu

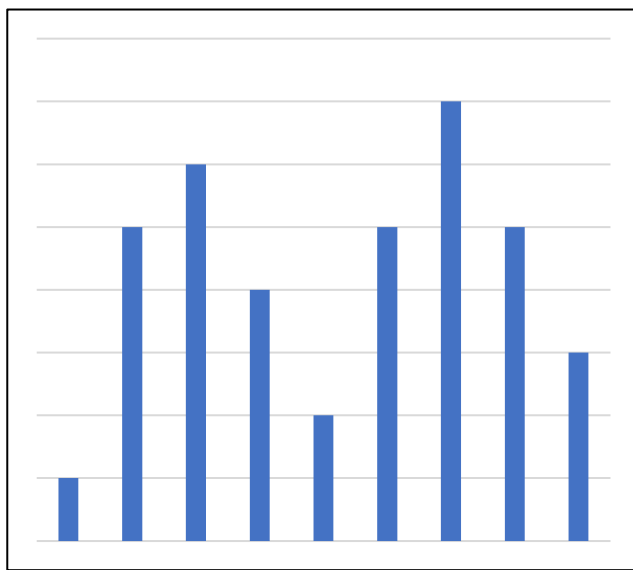
- Xu hướng: Trạng thái các mẫu hình dữ liệu xuất hiện thường xuyên.



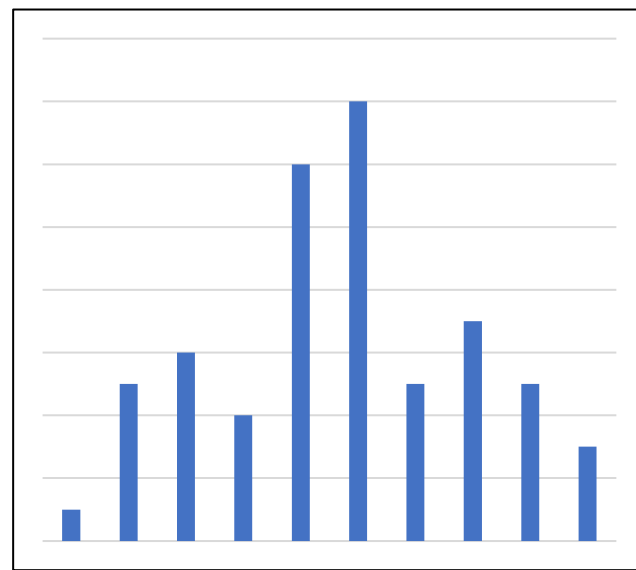
*Mẫu hình xiên
(skewness)*



*Mẫu đồng nhất
(uniform)*

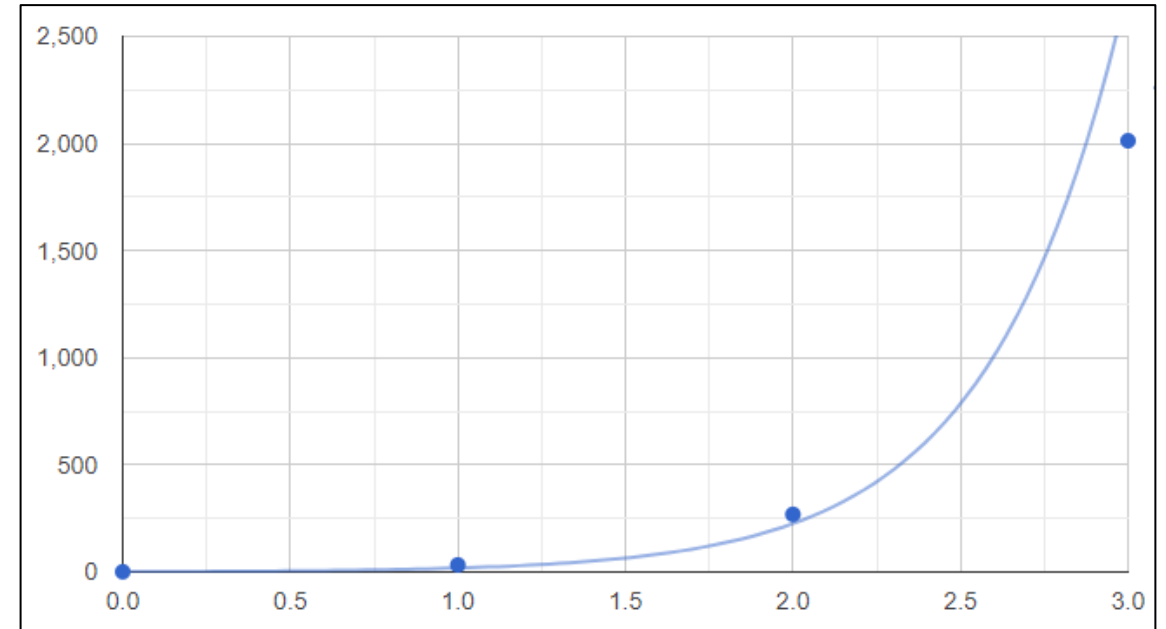
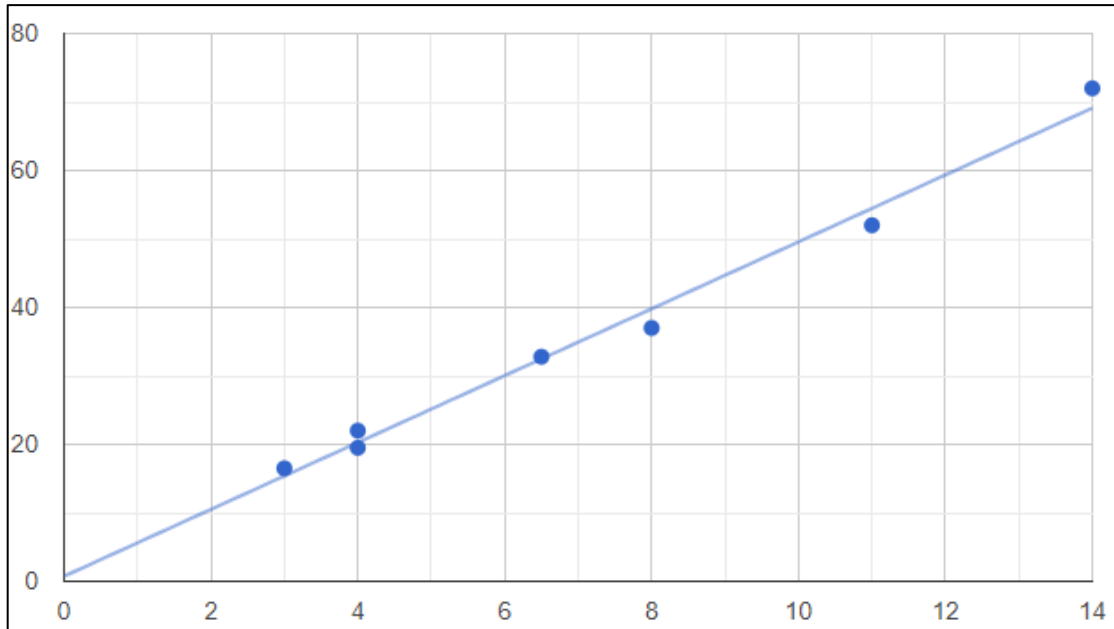


*Mẫu hình chuông
(bell-shaped)*



*Mẫu đầu & vai
(head & shoulders)*

- Khi biểu thị xu hướng bằng hàm số, đường biểu thị hàm số đó là đường xu hướng.



$$y = X\beta + \varepsilon$$

$$y = X\beta + \beta_0 I_m + \varepsilon$$

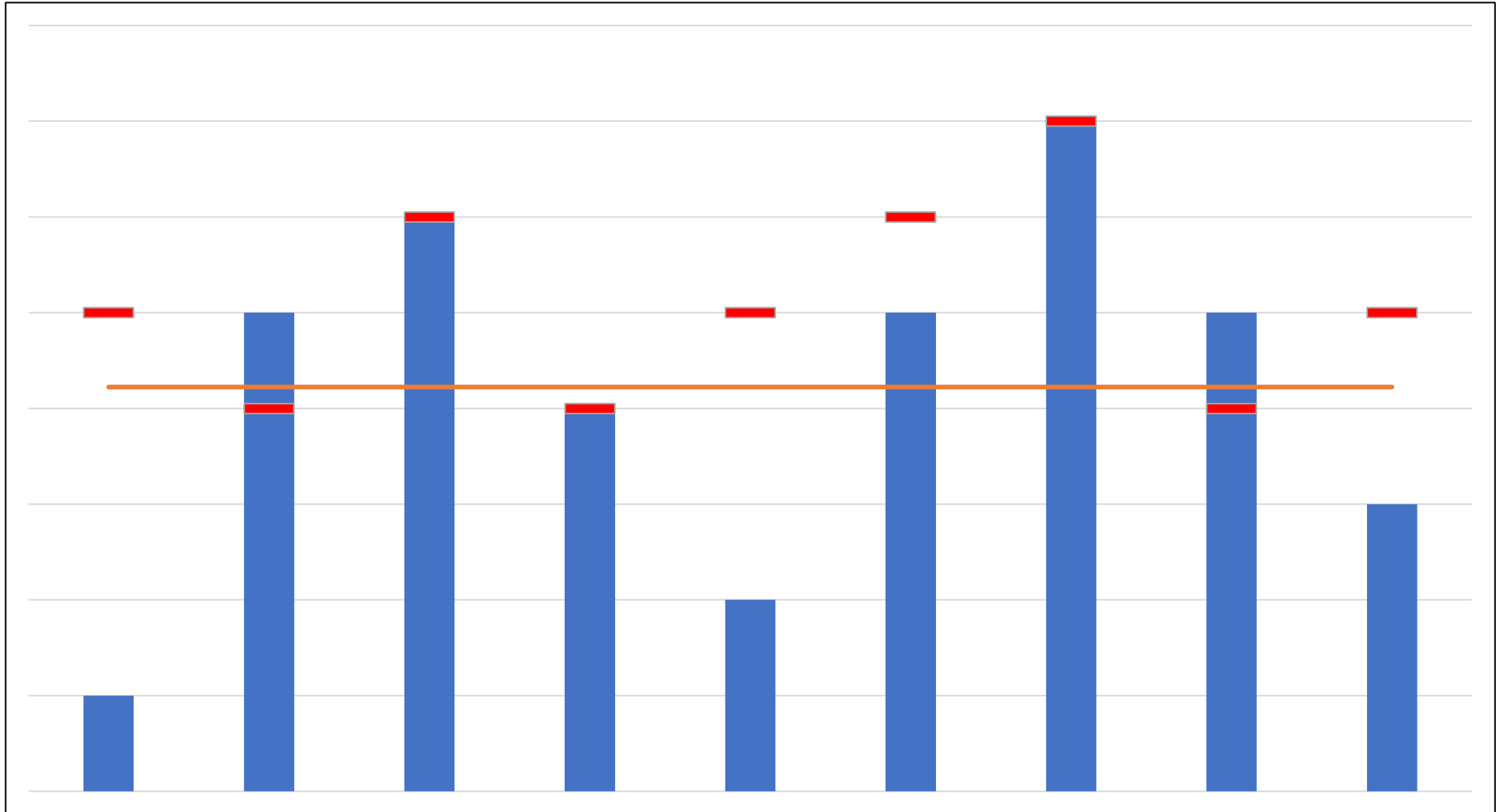
$$y = X\beta + \varepsilon$$

$$y = X\beta + \beta_0 I_m + \varepsilon$$

$$\beta = (X^T X)^{-1} X^T y$$

$$y = \beta x^2 + \beta x + \beta_0 + \varepsilon$$

- Đường mục tiêu đánh dấu giá trị kỳ vọng.



II. Thông tin dữ liệu và thao tác dữ liệu

- Lượng thông tin thường quyết định độ hữu ích dữ liệu.
- Xác định bằng hệ số entropy.

$$H(X) = \mathbb{E}_{X \sim P}[I(x)] = - \sum_{i=1}^m p(x_i) \log_2 p(x_i)$$

Shannon entropy

Thao tác lọc

1. Xác định điều kiện lọc
2. Tính kết quả điều kiện lọc
3. Lọc dữ liệu dựa trên kết quả phía trên

Thao tác nhóm

1. Xác định cột khóa
2. Xác định cột nhóm
3. Kiểm tra, lọc các giá trị hợp lệ
4. Thực hiện nhóm

Kết hợp dữ liệu

1. Chọn bảng kết hợp
2. Xác định chiều kết hợp, cột cơ sở
3. Xác định kiểu kết hợp: Inner, outer, left, right
4. Thực hiện kết hợp

III. Trực quan hóa trên Zeppelin

- Sử dụng các công cụ trực quan hóa có sẵn trong Zeppelin.