

Câu 1. Cho tập dữ liệu thu thập được của 2 tập giá trị của X và Y như sau:

X	5.1	4.9	4.6	5.7	5.4	5.1	5.1	5.3	5	7	6.4	4.5
Y	3.5	3	3.1	4.4	3.9	3.5	3.8	3.7	3.3	3.2	3.2	2.3

- a. Hãy xác định các giá trị trung bình, trung vị, mode của **X** và **Y**.
 - b. Vẽ biểu đồ Boxplot của **X** và **Y**.
 - c. Sử dụng phương pháp chuẩn hóa min-max để chuẩn hóa dữ liệu quan sát của **X** và **Y** về đoạn [1, 5].
 - d. Hãy làm trơn dữ liệu ban đầu của **X** và **Y** bằng phương pháp làm trơn trung bình (bin means), trong đó việc phân chia thùng theo chiều sâu (Equal-depth) với số bin là 3. Mô tả các bước thực hiện.
 - e. Xác định hệ số tương quan giữa **X** và **Y**.
- Câu 2.** Cho tập dữ liệu giao dịch sau:

ID	Items
T1	F, M, B, E
T2	O, F, E
T3	O, A, H, S
T4	D, B, F, E
T5	D, A, H, S
T6	O, M, E
T7	O, A, D, H, S

- a. Giả sử ngưỡng của độ hỗ trợ (minimum support) là 40%. Sử dụng thuật toán Apriori tìm tất cả các tập đối tượng thường xuyên.
- b. Giả sử ngưỡng của độ hỗ trợ là 40%, ngưỡng của độ tin cậy (minimum confidence) là 75%. Từ tập dữ liệu trên hãy xác định các luật kết hợp mạnh.