

**Câu 1.** Cho tập dữ liệu thu thập được giữa 2 biến **X** và **Y** như sau:

<b>X</b>	9.1	4.5	5.3	6.7	6.5	7.0	6.0	5.5	7.0	7.0	8.5	8.6
<b>Y</b>	8.5	5.5	6.0	7.5	8.5	6.0	6.5	6.8	9.0	7.2	8.0	8.5

- Hãy xác định các giá trị trung bình, trung vị, mode của **X** và **Y**.
- Vẽ biểu đồ Boxplot của **Y**.
- Hãy chuẩn hóa giá trị của **Y** về dạng chuẩn z-score.
- Hãy làm trơn dữ liệu ban đầu của **X** bằng phương pháp làm trơn thùng theo trung bình (bin means), trong đó việc phân chia thùng theo chiều rộng (Equal-width) với số thùng là 3.
- Xác định mối tương quan giữa **X** và **Y**.

**Câu 2.** Cho tập dữ liệu sau:

<b>ID</b>	<b>Items</b>
T1	I6, I1, I3
T2	I1, I2, I4, I5, I3
T3	I3, I2, I5
T4	I6, I7
T5	I1, I3, I2, I4, I5

- Giả sử ngưỡng của độ hỗ trợ (minimum support) là 60%. Sử dụng thuật toán Apriori tìm tất cả các tập đối tượng thường xuyên.
- Giả sử ngưỡng của độ hỗ trợ là 60%, ngưỡng của độ tin cậy (minimum confidence) là 75%. Từ tập dữ liệu trên hãy xác định các luật kết hợp mạnh.