**1.1.1. Ôn tập lý thuyết**

**+ Thống kê mô tả là gì? Nó khác gì với thống kê suy luận (inferential statistics)?**

**Trả lời:**

**+ Các thước đo thống kê mô tả chính (ví dụ: trung bình, trung vị, phương sai, độ lệch chuẩn) được sử dụng để làm gì? Trong trường hợp nào thì nên dùng trung vị thay vì trung bình?**

**+ Làm thế nào để xác định phân bố của một tập dữ liệu? Các loại phân bố phổ biến là gì (ví dụ: phân bố chuẩn, lệch trái, lệch phải)?**

**+ Độ lệch chuẩn và phạm vi (range) có ý nghĩa gì trong việc đánh giá sự phân tán của dữ liệu?**

**+ Sự khác biệt giữa các thước đo như Q1, Q2, Q3 trong biểu đồ hộp (boxplot) là gì?**

**Trả lời:**

Q1: Phản ánh mức độ phân chia của dữ liệu ở 25%, nếu tập trung nhiều dữ liệu ở điểm này thì dữ liệu sẽ bị lệch trái

Q2: Phản ánh mức độ phân chia của dữ liệu ở 50% (mean), nếu dữ liệu tập trung ở điểm này thì dữ liệu sẽ cân đối

Q3: Phản ảnh mức độ phân chia của dữ liệu ở 75%, nếu tập trung nhiều dữ liệu ở điểm này thì dữ liệu sẽ bị lệch phải.

**+ Làm thế nào để xử lý giá trị thiếu (missing values) trước khi tính toán các chỉ số thống kê mô tả?**

**+ Bạn có thể giải thích cách đọc và diễn giải một biểu đồ histogram hoặc boxplot từ dữ liệu thực tế không?**

**+ Khi gặp một tập dữ liệu có giá trị ngoại lai (outliers), bạn sẽ xử lý chúng như thế nào trước khi thực hiện thống kê mô tả?**