THỐNG KÊ MÁY TÍNH & ỨNG DỤNG

**Bài 1:** Cho dataset mpg.csv về mức tiêu hao nhiên liệu. Thông tin của dữ liệu như sau:

* mpg: Miles/(US) gallon số gallon xăng trên 1 dặm
* cylinders: số lượng xilanh
* displacement: Displacement/ dung tích xi lanh (cu.in / inches3)
* horsepower: công suất theo mã lực
* weight: trọng lượng (pound)
* acceleration: Gia tốc
* model\_year: Năm sản xuất
* origin: Xuất xứ
* name: Tên dòng xe

Hãy thực hiện:

#### (1.0đ) Đọc file dữ liệu và kiểm tra xem có dòng nào bị thiếu dữ liệu không? Nếu có, hãy xóa những dòng đó. Có dữ liệu nào bị trùng lặp không? Nếu có, xử lý thế nào?

#### (1.0đ) Tính giá trị trung bình và trung vị của các thuộc tính: 'cylinders', 'horsepower', 'weight'. Tìm miền giá trị (min, max) và phân vị bậc (Q1, Q3) của 'mpg'. Có xe nào có mức tiêu hao nhiên liệu đặc biệt cao hoặc thấp không?

#### (1.0đ) Tính mức tiêu hao nhiên liệu trung bình (mpg) của từng nhóm xe dựa trên 'origin'. Xe từ khu vực nào có xu hướng tiêu hao ít nhiên liệu nhất?

#### (1.0đ) Vẽ Histogram thể hiện sự phân bố của trọng lượng (weight) các xe. So sánh Boxplot của công suất (horsepower) giữa xe sản xuất trong thập niên 70 và 80. Xe từ thập niên nào có công suất cao hơn? Nhận xét xu hướng.

#### (1.0đ) Vẽ Scatter Plot thể hiện mối quan hệ giữa 'mpg' với các yếu tố: Số lượng xi-lanh (cylinders), Công suất (horsepower), Gia tốc (acceleration), Năm sản xuất (model\_year). Yếu tố nào có ảnh hưởng lớn nhất đến mức tiêu hao nhiên liệu?

**Bài 2:** Một nhà máy sản xuất vi mạch điện tử đang nghiên cứu về tỷ lệ lỗi trong sản phẩm, số lượng sản phẩm đạt tiêu chuẩn, cũng như thời gian hoàn thành mỗi lô hàng. Quá trình sản xuất được mô hình hóa bằng các phân phối xác suất phù hợp.

#### (a) (1,5 điểm) Trung bình mỗi lô hàng có 2 lỗi sản xuất. Giả sử số lỗi tuân theo phân phối Poisson. Tính xác suất một lô hàng nhiều hơn 3 lỗi.

#### (b) (1,5 điểm) Mỗi lô hàng gồm 20 vi mạch, và xác suất mỗi vi mạch đạt tiêu chuẩn là 95%. Tính xác suất một lô hàng có chính xác 18 vi mạch đạt tiêu chuẩn.

#### (c) (2 điểm) Thời gian hoàn thành một lô hàng tuân theo phân phối chuẩn, với trung bình 50 phút và độ lệch chuẩn 5 phút.

* Tính xác suất một lô hàng được hoàn thành trong vòng 45 phút.
* Để nằm trong top 5% lô hàng sản xuất nhanh nhất, cần hoàn thành trong thời gian bao nhiêu?