**Bài tập thực hành số 2**

**Mối quan hệ giữa các yếu tố lối sống, sinh hoạt và chỉ số sức khỏe với huyết áp tâm thu.**

**1. Bộ số liệu mẫu**

Bộ dữ liệu với 512 đối tượng với 10 biến độc lập và 1 biến phụ thuộc trong file **data2.csv** với các biến như dưới đây.

**Biến phụ thuộc:**

* HuyetApTamThu (Systolic Blood Pressure - Huyết áp tâm thu): Giá trị liên tục (mmHg)

**Biến độc lập:**

1. Tuoi (Age): Tuổi (năm, liên tục)
2. BMI (Body Mass Index): Chỉ số khối cơ thể (liên tục)
3. GioTapTheDuc (ExerciseHours): Số giờ tập thể dục/tuần (liên tục)
4. SoBuaAnCheBien (ProcessedMeals): Số bữa ăn chế biến sẵn/tuần (liên tục)
5. HutThuoc (Smoking): Tình trạng hút thuốc (0=Không, 1=Có - biến nhị phân)
6. UongRuou (Alcohol): Mức độ uống rượu (0=Không, 1=Ít, 2=Trung bình, 3=Nhiều - biến thứ bậc)
7. GioNgu (SleepHours): Số giờ ngủ trung bình/ngày (liên tục)
8. MucDoStress (StressLevel): Mức độ stress (thang điểm 1-10, liên tục)
9. TienSuBenhTim (HeartDiseaseHistory): Tiền sử bệnh tim mạch gia đình (0=Không, 1=Có - biến nhị phân)
10. Diabet (Diabetes): Mắc bệnh tiểu đường (0=Không, 1=Có - biến nhị phân)

**2. Nhiệm vụ**

1. Xây dựng mô hình hồi quy tuyến tính để dự đoán HuyetApTamThu dựa vào biến Tuoi. Phiên giải mô hình hồi qui theo các gợi ý sau

* Hệ số hồi quy của Tuoi là bao nhiêu? Nó có ý nghĩa thống kê không (dựa vào p-value)?
* Diễn giải ý nghĩa của hệ số hồi quy của Tuoi.
* Giá trị R-squared của mô hình là bao nhiêu? Nó nói lên điều gì?

2. Xây dựng mô hình hồi quy tuyến tính để dự đoán HuyetApTamThu dựa trên *tất cả* các biến độc lập đã cho.

* Biến độc lập nào có ý nghĩa thống kê nhất trong mô hình này?
* So sánh R-squared của mô hình đơn biến và đa biến. Mô hình nào giải thích được nhiều phương sai của HuyetApTamThu hơn?
* Giải thích ý nghĩa của hệ số hồi quy cho biến BMI và HutThuoc (nếu có ý nghĩa thống kê).

3. Kiểm tra giả định của mô hình hồi quy đa biến

* Kiểm tra xem mối quan hệ giữa các biến độc lập liên tục và biến phụ thuộc có tuyến tính hay không. Sử dụng biểu đồ phân tán hoặc biểu đồ phần dư.
* Kiểm tra xem phần dư của mô hình có phân phối gần chuẩn hay không
* Kiểm tra xem phương sai của phần dư có đồng nhất trên tất cả các mức của biến độc lập hoặc giá trị dự đoán hay không
* Kiểm tra mức độ đa cộng tuyến giữa các biến độc lập và đưa ra phương án xử lý nếu cần

4. Trình bày kết quả của mô hình đa biến cuối cùng (sau khi đã xử lý đa cộng tuyến nếu cần).