

Câu 2: (50đ) – Bài tập thực hành

Dữ liệu về mức độ sử dụng xe buýt tại 40 thành phố của Mỹ ghi nhận trong cùng 1 năm có trong file **P09-Data.xls**, trong đó bao gồm các biến cụ thể sau:

- BUSTRAVL: mức độ sử dụng xe buýt để di chuyển ở mỗi thành phố (ngàn hành khách/giờ)
- FARE: giá vé xe buýt (tính bằng USD)
- GASPRICE: giá một nhiên liệu (tính bằng USD/ga lông)
- INCOME: thu nhập bình quân đầu người (tính bằng USD)
- POP: dân số của từng thành phố (ngàn người)
- DENSITY: mật độ dân số (người/dặm vuông)
- LANDAREA: diện tích thành phố (dặm vuông)

Giả sử các lý thuyết kinh tế cho bạn biết những biến trên FARE, GASPRICE, INCOME, POP, DENSITY, LANDAREA sẽ ảnh hưởng đến BUSTRAVL theo dạng hàm tuyến tính.

Anh/chị hãy thực hiện các yêu cầu sau đây:

- a. Anh chị kỳ vọng như thế nào về dấu của các hệ số hồi quy đứng trước các biến độc lập?
- b. Vẽ đồ thị phân tán scatter thể hiện quan hệ giữa các biến độc lập với biến phụ thuộc? Lập ma trận tương quan giữa các biến độc lập và biến phụ thuộc và nhận xét từ kết quả?
- c. Anh chị hãy ước lượng mô hình không giới hạn (U), giải thích và đưa ra các kết luận ngắn gọn từ kết quả ước lượng?

Mô hình (U): $BUSTRAVL = f(FARE, INCOME, POP, DENSITY, LANDAREA) + u$

- d. Anh chị hãy xác định mô hình giới hạn (R) tốt nhất có thể theo anh/chị. Lưu ý, việc loại bỏ các biến không phù hợp phải được kiểm định và chứng minh.
- e. Anh chị hãy lấy Log cơ số e (với $e = 2.7182$) của tất cả các biến trong mô hình R bao gồm cả biến độc lập và phụ thuộc mà anh/chị đã chọn ở câu d. Sau đó ước lượng lại mô hình hồi quy mới với các biến ln. So sánh 2 mô hình, anh/chị chọn mô hình nào? Vì sao?
- f. **(Không bắt buộc)** Sau khi đã có mô hình (R) ở Câu d. Anh/chị hãy ứng dụng Kiểm định Lagrange để chứng minh việc loại các biến ra khỏi mô hình (U) là có cơ sở?

---HẾT---