KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
227 Nguyễn Văn Cừ, Phường 4, Quận 5, TP.HCM
Điện thoại: (08) 38.354.266 – Fax: (08) 38.350.096

CSC12110 – PHÂN TÍCH DỮ LIỆU ỨNG DỤNG

BÀI TẬP LÝ THUYẾT

BTLT 4: HỒI QUY

I. Thông tin chung

Mã số bài tập:	BTLT 4
Thời lượng dự kiến:	3 tiếng
Deadline nộp bài:	18/12/2023 23:00:00
Hình thức:	Bài tập cá nhân
Hình thức nộp bài:	Nộp qua Moodle môn học
GV phụ trách:	Vũ Thị Mỹ Hằng
Thông tin liên lạc với GV:	vtmhang@fit.hcmus.edu.vn

II. Mô tả bài tập

Câu hỏi 1: Trường ĐH KHTN đã thực hiện điều tra ngẫu nhiên một mẫu gồm các cựu sinh viên để hiểu về sự liên hệ giữa điểm số của sinh viên ra trường và mức lương khởi điểm. Kiểm tra sự tương quan giữa hai biến số.

Dữ liệu: Tập dữ liệu bao gồm điểm GPA khi ra trường của sinh viên và mức lương khởi điểm tính bằng USD.

a) Cài đặt code đánh giá mức độ tương quan thông qua mô hình hồi quy. Cần giải thích rõ các thông số thu được từ kết quả.
b) Dự đoán mức lương khởi điểm cho sinh viên có điểm trung bình khi học đại học là 8.0.

Câu hỏi 2: Sử dụng hồi quy tuyến tính để dự đoán Tổng sản phẩm quốc nội (GDP) của một quốc gia dựa trên các yếu tố như đầu tư công cộng, tiêu dùng cá nhân, và tỷ lệ thất nghiệp.

Dữ liệu: Tập dữ liệu bao gồm GDP hàng năm, đầu tư công cộng, tiêu dùng cá nhân, và tỷ lệ thất nghiệp cho một quốc gia trong vòng 10 năm qua.

a) Xây dựng mô hình hồi quy tuyến tính sử dụng đầu tư công cộng, tiêu dùng cá nhân, và tỷ lệ thất nghiệp làm biến độc lập.
b) Đánh giá mô hình dựa trên các chỉ số như R-squared, F-statistic.
c) Đánh giá mức độ phụ thuộc của GDP vào từng biến dựa trên các chỉ số p-value cho từng biến.
d) Dự đoán GDP cho năm tiếp theo dựa trên các ước lượng từ mô hình.

Câu hỏi 3: Sử dụng lại dataset ở BTLT 2 (E2.3), dùng phương pháp hồi quy điền vào missing value cho cột overall scores.