**Bài tập thực hành 4:**

**Hồi quy Poisson dự đoán số lần có ý định tự tử**

Nghiên cứu các yếu tố liên quan đến số lần có ý định tự tử ở nhóm đối tượng trẻ (18-25 tuổi), đặc biệt nhấn mạnh mối liên hệ với trầm cảm.

**1. Bộ số liệu mẫu**

Chúng ta sẽ sử dụng bộ dữ liệu phân tích mô hình hồi quy logistics trước đó (**data4.csv**) và thêm một biến SoLanTuTu (số lần có ý định tự tử). Bộ dữ liệu sẽ có 350 đối tượng ở độ tuổi 18-25 với 10 biến độc lập ban đầu, biến phụ thuộc TramCam (nhị phân), và một biến phụ thuộc mới là SoLanTuTu (biến đếm).

**Biến phụ thuộc mới:**

1. SoLanTuTu: Số lần có ý định tự tử trong 6 tháng gần nhất (biến đếm, giá trị 0,1,2,...)

**Các biến khác (như bài tập trước):**

1. TramCam: Tình trạng trầm cảm (0 = Không trầm cảm, 1 = Có trầm cảm - biến nhị phân)
2. Tuoi : Tuổi (năm, liên tục)
3. GioHoc : Số giờ học/làm việc/tuần (liên tục)
4. GioNgu : Số giờ ngủ trung bình/ngày (liên tục)
5. MoiQuanHeXaHoi : Điểm số mối quan hệ xã hội (thang điểm 1-10, liên tục)
6. TienSuTramCamGiaDinh : Tiền sử trầm cảm trong gia đình (0 = Không, 1 = Có - biến nhị phân)
7. ApLucHocTap : Mức độ áp lực học tập (thang điểm 1-10, liên tục)
8. HoatDongTheChat :Mức độ hoạt động thể chất (0 = Ít, 1 = Trung bình, 2 = Nhiều - biến thứ bậc)
9. SuDungThucPham : Mức độ tiêu thụ thực phẩm không lành mạnh (0 = Ít, 1 = Trung bình, 2 = Nhiều - biến thứ bậc)
10. ThuNhapGiaDinh : Thu nhập gia đình (triệu VND/tháng, liên tục)
11. SuDungMangXaHoi : Số giờ sử dụng mạng xã hội/ngày (liên tục)

**Nhiệm vụ:**

1. Xây dựng mô hình hồi quy Poisson để dự đoán SoLanTuTu chỉ dựa vào biến TramCam

* Hệ số hồi quy của TramCam1 là bao nhiêu? Nó có ý nghĩa thống kê không (dựa vào p-value)?
* Làm thế nào để chuyển hệ số hồi quy này thành Tỷ số tỷ lệ (Rate Ratio - RR)? Diễn giải ý nghĩa của RR này.
* Giá trị AIC của mô hình là bao nhiêu?

1. Xây dựng mô hình hồi quy Poisson để dự đoán SoLanTuTu dựa trên *tất cả* các biến độc lập (bao gồm cả TramCam).

* Những biến độc lập nào có ý nghĩa thống kê trong mô hình này (p-value < 0.05)?
* So sánh AIC của mô hình đơn biến và đa biến. Mô hình nào "tốt hơn" dựa trên tiêu chí này?
* Giải thích ý nghĩa của Rate Ratio cho biến TramCam1 và GioNgu (nếu có ý nghĩa thống kê).
* Rate Ratio của ApLucHocTap là bao nhiêu?

1. Kiểm tra xem mô hình có bị overdispersion hay không

* Giá trị Residual Deviance / DF của mô hình là bao nhiêu? Bạn có nhận thấy dấu hiệu của overdispersion không?
* Kết quả kiểm định cho bạn biết điều gì.

1. Tương tự như Logistic, hồi quy Poisson giả định mối quan hệ tuyến tính giữa biến độc lập liên tục và log của tỷ lệ trung bình (log-rate)

* Dựa trên biểu đồ LOESS, mối quan hệ giữa biến độc lập liên tục (ví dụ ApLucHocTap) và log-rate có vẻ tuyến tính không?
* Nếu không tuyến tính, bạn sẽ đề xuất cách xử lý nào?

1. Kiểm tra mức độ đa cộng tuyến giữa các biến độc lập

* Giá trị VIF của từng biến là bao nhiêu?
* Có biến nào có VIF cao (thường > 5 hoặc > 10) không? Nếu có, đó là biến nào?
* Nếu có biến có VIF cao, bạn đề xuất cách xử lý nào?