

Đề bài: KHẢO SÁT DỮ LIỆU VỀ 3 LOÀI CHIM CÁNH CỤT

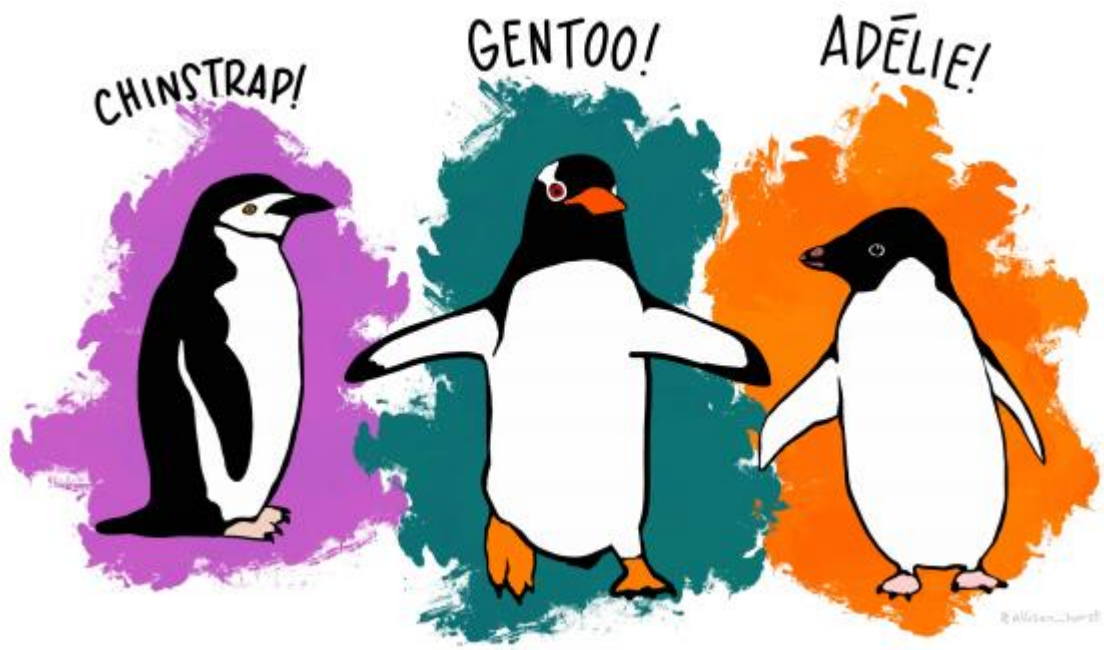
Dữ liệu được cung cấp trong file *penguins-data.csv* được khảo sát các số đo của 3 loài chim cánh cụt trên 3 đảo (Biscoe, Dream, Torgersen). Yêu cầu

1. **(2 điểm)** Viết mã để đọc dữ liệu và tìm hiểu thông tin về dữ liệu:
 - a. Dữ liệu có bao nhiêu chim cánh cụt. Số lượng mỗi loại
 - b. Thể hiện các thông số thống kê cơ bản của từng thuộc tính thông qua hàm **describe()**. Dựa vào kết quả bạn hãy đưa ra bình luận trên thuộc tính trọng lượng (`body_mass_g`)
2. **(4 điểm)** Khảo sát thuộc tính trọng lượng (`body_mass_g`)
 - a. Vẽ biểu đồ cho biết phân phối của trọng lượng của những chim cánh cụt trong bộ dữ liệu
 - b. Vẽ biểu đồ so sánh trọng lượng của ba loại chim cánh cụt
 - c. Theo bạn thuộc tính định lượng nào có khả năng có mối tương quan tuyến tính với thuộc tính định lượng
 - d. Lựa chọn thuộc tính có mối tương quan tuyến tính cao nhất với thuộc tính trọng lượng. Bạn hãy xây dựng mô hình hồi quy tuyến tính đơn giản để dự đoán trọng lượng chim cánh cụt dựa trên thuộc tính này. Viết hàm hồi quy tìm được.
3. **(4 điểm)** Giả sử bộ dữ liệu được cung cấp là một mẫu thống kê. Bạn hãy thực hiện một số ước lượng sau liên quan đến loài chim cánh cụt Gentoo
 - a. Ước lượng tỷ lệ loài chim cánh cụt Gentoo.
 - b. Ước lượng trọng lượng của của loài chim cánh cụt Gentoo.

Lưu ý:

- Bài thi lưu theo định dạng **<hoten>_TH.ipynb**
- Phần đầu file bài làm ghi thông tin: mã sinh viên, họ tên
- Các câu trả lời nhận xét, bình luận gõ trực tiếp vào file bài làm

----- Hết -----



Body Part of Penguin

