

TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP ĐỀ MINH HỌA THI GIỮA KÌ 01

THÀNH PHỐ HỒ CHÍ MINH

Môn thi: Thống Kê Máy Tính Và Ứng Dụng

Lớp: HTTT18

KHOA CÔNG NGHỆ THÔNG TIN

Ngày thi: 2024

Thời gian: phút

Họ và tên thí sinh; MSSV:

Đề bài: DỮ LIỆU SINH VIÊN

Dữ liệu được cung cấp trong file StudentSurvey.csv được khảo sát. Yêu cầu

1. **(1 điểm)** Hiện thị 5 dòng dữ liệu
(1 điểm) Xây dựng bảng phân phối tần số thể hiện số lượng sinh viên mỗi năm, bỏ qua các dòng bị thiếu dữ liệu năm.
2. Dựa vào dữ liệu bạn hãy cho biết:
 - a. **(2 điểm)** Tỷ lệ sinh viên đeo khuyên?
 - b. **(1 điểm)** Giữa sinh viên nam và nữ tỷ lệ đeo khuyên của ai cao hơn?
3. **(1.5 điểm)** Vẽ đồ thị histogram của ba thuộc tính nhịp tim (Pulse), số giờ tập thể thao (Exercise), số khuyên đeo (Piercings).
(0.5 điểm) Hãy cho biết phân phối của ba thuộc tính trên phân phối nào lệch trái, lệch phải hoặc đối xứng.
4. **(3 điểm)** Sử dụng hàm sample() trên DataFrame. Bạn hãy:
 - a. **(1 điểm)** Lấy ngẫu nhiên một mẫu từ bộ dữ liệu trên với cỡ mẫu là 50. Tính tỷ lệ nữ mẫu trên. Tính sai số so với giá trị tỷ lệ từ bộ dữ liệu ban đầu.
 - b. **(1 điểm)** Hãy thực hiện lại câu trên 100 lần với cùng cỡ mẫu là 50, với mỗi mẫu hãy tính giá trị tỷ lệ nữ của mẫu và vẽ histogram của 100 giá trị trung bình này
 - c. **(1 điểm)** Lập lại câu b và thử nghiệm với các cỡ mẫu khác nhau. Bạn có nhận thấy có gì khác biệt không? Nếu có hãy giải thích

Lưu ý:

Bài thi lưu theo định dạng <stt>_<hoten>_GK.ipynb

Phần đầu file bài làm ghi thông tin: mã sinh viên, họ tên, số máy.

Các câu trả lời nhận xét, bình luận gõ trực tiếp vào file bài làm

-----**Hết**-----

StudentSurvey

Year	Year in school: First Year, Sophomore, Junior, or Senior
Gender	Student's gender: F or M
Smoke	Smoker? No or Yes
Award	Preferred award: Academy, Nobel, or Olympic
Higher SAT	Which SAT is higher? Math or Verbal
Exercise	Hours of exercise per week
TV	Hours of TV viewing per week
Height	Height (in inches)
Weight	Weight (in pounds)
Siblings	Number of siblings
Birth Order	Birth order, 1 oldest, 2 second oldest, etc.
VerbalSAT	Verbal SAT score
MathSAT	Math SAT score
SAT	Combined Verbal + Math SAT
GPA	College grade point average
Pulse	Pulse rate (beats per minute)
Pulseing	Number of body Pulse