

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP
THÀNH PHỐ HỒ CHÍ MINH**

KHOA CÔNG NGHỆ THÔNG TIN

ĐỀ MINH HỌA THI GIỮA KÌ 04

Môn thi: Thống kê máy tính và ứng dụng

Lớp: DHTH19

Ngày thi: 2024

Thời gian: 60 phút

Họ và tên thí sinh; MSSV:

Câu 1 (6 điểm)

Dataset `acs12.csv` khảo sát cộng đồng người Mỹ năm 2012:

- **income:** thu nhập hàng năm
- **employment:** tình trạng việc làm
- **hrs_work:** số giờ làm việc mỗi tuần
- **race:** chủng tộc
- **age:** tuổi
- **gender:** giới tính
- **citizen:** có phải là cư dân của Mỹ không?
- **time_to_work:** thời gian đến sở làm (phút)
- **lang:** ngôn ngữ sử dụng ở nhà
- **married:** tình trạng hôn nhân
- **edu:** trình độ học vấn
- **disability:** có phải là người khuyết tật không?

Whether the person is disabled.

Bạn hãy trả lời câu hỏi sau:

a. Đọc bộ dữ liệu và cho biết một số thông tin sau:

- Kích thước?
- Hiển thị 7 dòng dữ liệu cuối cùng

b. Với mỗi thuộc tính định tính cho biết các giá trị phân biệt

c. Tính tỉ lệ mỗi trình độ học vấn trong bộ dữ liệu.

- **NA**: không có dữ liệu (not available)
- **hs or lower**: trình độ THPT trở xuống
- **college**: trình độ Đại học
- **grad**: trình độ sau Đại học (Master or Doctor)

d. So sánh trình độ tốt nghiệp sau Đại Học giữa các chủng tộc (**'race'**) khác nhau

e. Vẽ pie-chart thể hiện tỉ lệ phần trăm các chủng tộc (**'race'**) trong khảo sát

f. Vẽ đồ thị khảo sát sự ảnh hưởng giữa số giờ làm việc (**'hrs_work'**) đến thu nhập (**'income'**)

Câu 2 (4 điểm)

Giả sử bạn có một túi đựng bóng với tỷ lệ bóng đỏ là $p=0.4$ và bóng xanh là $1-p=0.6$

Lấy ngẫu nhiên 10 bóng từ túi

- Viết mã python để tính xác suất để có đúng 4 bóng đỏ
- Viết mã python để tính xác suất để có đúng 2 bóng xanh
- Tổng quát gọi X là số bóng xanh lấy được. Vẽ đồ thị thể hiện xác suất các giá trị X có thể nhận được
- Tổng quát gọi X là số bóng xanh lấy được. Vẽ đồ thị thể hiện xác suất tích lũy các giá trị X có thể nhận được

-----**Hết**-----

Lưu ý:

- Bài thi lưu theo định dạng **<msv>_<hoten>_GK.ipynb** và **<msv>_<hoten>_GK.html**
- 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