

| **TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP**  **THÀNH PHỐ HỒ CHÍ MINH**    **KHOA CÔNG NGHỆ THÔNG TIN** |  |  |
| --- | --- | --- |

Họ tên: ……………………………Mã SV:…………………Số máy:………

Sinh viên nộp bài một file duy nhất ***HoTen\_MaSV.\****

Cho dữ liệu dataKHXH.csv chứa điểm thi THPT quốc gia năm 2023 của các thí sinh tham gia khối Khoa Học Xã Hội.

Sử dụng hàm ***data.sample(n, random\_state=15)*** để lấy ngẫu nhiên n phần từ data

1. ***(4đ)***
   1. Đọc dữ liệu và cho biết có bao nhiêu thí sinh trong bảng dữ liệu. Tính điểm trung bình, độ lệch chuẩn, phương sai của điểm môn Ngữ Văn ("ngu\_van") (0.5đ)
   2. Chọn ngẫu nhiên 25 mẫu, sử dụng 25 mẫu này để ước lượng điểm trung bình của môn Ngữ Văn (**sử dụng độ lệch chuẩn tính được bên trên làm độ lệch chuẩn quần thể**) với độ tin cậy là 95%. (3đ)
   3. Từ kết quả ước lượng bên trên hãy cho nhận xét về giá trị trung bình ước lượng và giá trị thật tính từ dữ liệu. (0.5đ)
2. ***(4đ)*** Lấy ngẫu nhiên 1000 mẫu dữ liệu từ dữ liệu gốc, có nhận xét cho rằng: "**Có từ** 75% thí sinh khối KHXH có điểm Toán trên trung bình". Bạn có thể đưa ra kết luận gì nhận xét trên dựa vào 1000 mẫu vừa lấy được với mức ý nghĩa 5%? (4đ)
3. ***(2đ)*** Khảo sát điểm trung bình các bài lab của một môn học **X** và điểm thi cuối kỳ **Y** trong một môn học thực hành tại trường. Khảo sát ngẫu nhiên 8 sinh viên, ta thu được bảng số liệu sau:

| TB lab | 7.5 | 5.5 | 3.0 | 8.0 | 9.0 | 1.0 | 6.0 | 6.0 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Cuối kỳ | 8 | 6.5 | 4 | 7.5 | 9.5 | 3.0 | 5.0 | 6.0 |

* 1. Dựa vào bảng dữ liệu trên cho biết có xây dựng được mô hình hồi quy hay không? Nếu có hãy xây dựng mô hình hồi quy để **dự đoán điểm trung bình các bài lab dựa theo điểm thi cuối kỳ**. (1đ)
  2. Dự đoán điểm cuối kỳ của một sinh viên có điểm TB lab là 6.2. (1đ)

------------------------- Hết ------------------------------

*Sinh viên* ***tuyệt đối*** *không trao đổi bài, các bài giống nhau bị điểm 0 Thực hành*